

Proceedings of the Symposium on Assessment and Learner Outcomes

Rutherford House, Victoria University, Wellington, New Zealand
1-3 September 2011

Edited by Mimi Hodis and Susan Kaiser
Jessie Hetherington Centre for Educational Research
Victoria University
Wellington, New Zealand

Publication date: October 2012
ISBN: 978-0-475-12398-5



Table of Contents

INTRODUCTION

Editorial Board	ii
Editorial Reviewers	ii

SECTION 1: CONFERENCE PAPERS

External and Internal Moderation: The Other Side of the Story	2
<i>Tahera Afrin</i>	
The Assessment of Literacy Capability in Adults: Tensions and Complexities.....	15
<i>Helen Anderson and Lisa Maurice-Takerei</i>	
Formative Learning: A Description of an Assessment Framework in Initial Teacher Education.....	25
<i>Ann Balcombe, Liz Everiss, and Margaret Brennan</i>	
The Future of Secondary School Examinations: The Use of Technology in the New Zealand Context	41
<i>Stephen Bargh</i>	
Exploring Four Self-Assessment Strategies to Sustain Learning in Tertiary Education	57
<i>Roseanna Bourke</i>	
How do Students use Self-assessment to Learn in a Tertiary Institution?	70
<i>Roseanna Bourke and Carolyn Tait</i>	
Two Uses of Value-Added Modelling in New Zealand Secondary Schools	81
<i>Therese M. Boustead</i>	
Assessing Group Work in Student Industry Projects: Is Fairness Achievable?.....	92
<i>Jill Clark and Trish Baker</i>	
Popular Culture, Overall Teacher Judgments and National Standards.....	107
<i>John Dickie and Mary Jane Shuker</i>	
Promoting Information Literacy Development by Assessing the Research Process in First-year Tertiary Assessments.....	121
<i>Angela Feekery & Lisa Emerson</i>	
Reducing exclusions and improving attainment – how can that be? Evidence from a United Kingdom secondary school.....	137
<i>Gwen Gilmore</i>	
Enhancing Learner Outcomes Through the Use of Rubrics in the Summative Evaluation Process	152
<i>Trudy Harris, Dorothy Spiller, Michele Schoenberger-Orgad and Cheryl Cockburn-Wootten</i>	
Design Experiments for Assessment of “21st Century” Learning Outcomes	164
<i>Rosemary Hipkins</i>	

An Alignment of the Trends in International Mathematics and Science Survey Mathematics Scale with the 'By the End of Year 5' National Standard in Mathematics ..	174
<i>Elliot Lawes</i>	
Beyond Assessment: Assuring Student Learning in Higher Education	187
<i>Victoria J. Mabin and Stephen J. Marshall</i>	
Schools Out – Gone to the Beach: How Some Year 11 New Zealand High School Students Perceive Study Leave	204
<i>Peter Sanders</i>	
The ORBIT Online Analysis, Profiling and Reporting Tool.....	221
<i>Peter Tait, Michael Davison, John Hattie, Wendy Kofoed</i>	
The Benefits of Effective Behaviour Screening for Children in Schools and Kindergartens	231
<i>Gaye Tyler-Merrick and John Church</i>	
Reading Achievement over the Transition to Secondary School: Profiling a Problem	242
<i>Aaron Wilson, Stuart McNaughton, and Mei Kuin Lai</i>	
New Zealand's First Tertiary High School: Students' Perspectives of Various Aspects of Their Education and Experiences in School	258
<i>Colleen Young</i>	

SECTION 2: KEYNOTE SPEAKERS

Assessment in a Changing World	275
<i>Eva Baker</i>	
National Education Standards for New Zealand: A Research Agenda	285
<i>John Hattie</i>	
Researching Quality and Equity in Assessment for Improved Learner Outcomes	296
<i>Val Klenowski</i>	
The NCEA as an Exemplar of Research into Practice: Motivation, Achievement and Assessment Design	307
<i>Luanna H. Meyer</i>	
Full List of Symposium Contributors.....	313

Introduction

The Symposium that resulted in these published papers was co-sponsored by Victoria University, the University of Auckland, the Ministry of Education, and the New Zealand Qualifications Authority. Special appreciation is due to the conference organising committee comprising John Hattie, Rawiri Hindle, Mimi Hodis, Jenny Horsley, Michael Johnston, Kiritina Johnstone, Mary Klaver, Mei Kuin Lai, Luanna H. Meyer, Boaz Shulruf, and Louise Starkey. Thank you especially to Graeme Aitken, Bali Haque, Lisa Rodgers, and Dugald Scott, for their facilitation of support from their respective institutions and agencies for the Symposium itself. Special thanks are also due to Pam Ritchie who was the conference administrator and provided key support to the organisers and participants.

The individual papers published in these online Proceedings from the Symposium represent those submitted by their authors following the symposium. Each of these published papers was quality assured through an independent, anonymous peer review process and editorial review by both New Zealand and international experts in the relevant specialised areas of assessment. Not all papers presented at the Symposium are published in these proceedings, and interested persons may consult the full list of papers at the end of the Proceedings and contact authors directly for materials of interest. We owe a debt of gratitude to many who contributed their expertise to the Symposium through presentations and/or reviewing papers as part of our quality assurance process. Thank you.

Conference Organising Committee
October 2012

Editorial Board

Rawiri Hindle
Victoria University of Wellington
New Zealand

Mimi Hodis
Victoria University of Wellington
New Zealand

Jenny Horsley
Victoria University of Wellington
New Zealand

Luanna H. Meyer
Victoria University of Wellington
New Zealand

Boaz Shulruf
Auckland University
New Zealand

Louise Starkey
Victoria University of Wellington
New Zealand

Editorial Reviewers

Tahera Afrin
ICL Business School
Auckland
New Zealand

Dayle Anderson
Victoria University of Wellington
New Zealand

Helen Anderson
Manakau Institute of Technology
Auckland
New Zealand

Roseanna Bourke
Victoria University of Wellington
New Zealand

Gavin Brown
Auckland University
New Zealand

Sue Cherrington
Victoria University of Wellington
New Zealand

Vijaya Dharan
Victoria University of Wellington
New Zealand

Stephanie Doyle
Victoria University of Wellington
New Zealand

Michael Drake
Victoria University of Wellington
New Zealand

Ian M. Evans
Massey University
Wellington
New Zealand

Lester Flockton
University of Otago
Dunedin
New Zealand

Cedric Hall
Victoria University of Wellington
New Zealand

Roger Harvey
Victoria University of Wellington
New Zealand

Trudy Harris
Waikato University
New Zealand

Eva Heinrich
Massey University
New Zealand

Rose Hipkins
New Zealand Council for Educational
Research
Wellington
New Zealand

Flaviu Hodis

Victoria University of Wellington
New Zealand

Mimi Hodis

Victoria University of Wellington
New Zealand

Anne Hynds

Victoria University of Wellington
New Zealand

Michael Johnston

Victoria University of Wellington
New Zealand

Liz Jones

Victoria University of Wellington
New Zealand

Ella Kahu

Massey University
Palmerston North
New Zealand

Mei Kuin Lai

Auckland University
New Zealand

Kathleen Liberty

Canterbury University
New Zealand

David Lillis

New Zealand Qualifications Authority
Wellington
New Zealand

Stephen Marshall

Victoria University of Wellington
New Zealand

Azra Moeed

Victoria University of Wellington
New Zealand

Eleanor Rowe

Auckland University
New Zealand

Catherine Savage

Te Tapuae o Rehua
Christchurch
New Zealand

Boaz Shulruf

Auckland University
New Zealand

Louise Starkey

Victoria University of Wellington
New Zealand

Alison Stephenson

Victoria University of Wellington
New Zealand

Rob Strathdee

Victoria University of Wellington
New Zealand

Carolyn Tait

Victoria University of Wellington
New Zealand

SECTION 1:
Conference Papers

External and Internal Moderation: The Other Side of the Story

Tahera Afrin

Early Childhood Education (ECE), ICL Business School

Email: tahera@icl.ac.nz

Abstract

Tertiary Education Organisations (TEOs) that offer programmes including unit standards set by the New Zealand Qualifications Authority (NZQA) are required to participate in the national external moderation system. In spite of conducting regular internal moderation, TEOs often fail when internally moderated unit standard materials are sent for external moderation. This paper shares the experience of a non-university training provider and analyses it further to identify factors that lead to ineffective moderation. The initial discussion concerns the two types of moderation that TEOs are involved in: internal and external moderation. Later it focuses on the issues and challenges related to these processes. The background research covers nine unit standards that count towards the National Certificate in Early Childhood Education and Care (level 5), taught and moderated by the teaching staff of a Private Training Establishment (PTE) in Auckland. The Content Analysis Method is used to analyse the moderation reports and moderation meeting minutes. The result reveals conceptual, realistic, preparatory, ethical, organisational and structural concerns both at the pre-moderation and post-moderation stages. The paper also includes recommendations for TEOs, lecturers working for TEOs and NZQA to change their approach towards moderation. The findings of this research will assist all relevant stakeholders to implement effective internal and external moderation.

Keywords: *Moderation, Tertiary Education Organisations, NZQA.*

In relation to the unit-standard based qualifications system, moderation is a broad term that covers activities which help to ensure that there is a uniform interpretation and application of standards (New Zealand Qualifications Authority (NZQA), 1992). Within educational institutes, moderation is the process of sharing expectations and understanding of standards to improve consistency in teachers' decisions about students' learning (TKI, 2011). The three purposes of moderation as summarised by *Te Kete Ipurangi* (TKI) are to make reliable, valid, evidence-based decisions, to make consistent decisions over time, and to support the assessment of learning. These purposes indicate that moderation is to do with students' learning. Therefore, the moderation process has a fundamental stage that begins with the planning of learning. At the planning stage, certain goals to be achieved by the learners are determined. Teaching-learning practices in the classroom are the overlooked part of moderation that involves educators and learners. The following step is for learners to complete pre-designed assessment tasks. The teachers (markers) mark their efforts and make decisions about the competency of each learner. The noticeable part of moderation starts at this point where samples of marked assessments are gathered and given to another staff member (moderator) to check that the assessment tasks in the completed marked samples reflect the planning, that the samples are marked completely and uniformly and also whether the marker's decisions about the competency of the learners are valid or not. The last step of moderation is for the teachers to get feedback from the moderators so that they make changes in the assessment tasks or in their marking practice if required.

Non-university Tertiary Education Organisations (TEOs) participate in these moderation activities in partnership with the NZQA to ensure that valid, fair, accurate and consistent internal assessments are made (NZQA, 2011b). The moderation activities at the TEOs can be discussed in four different categories: i) pre-moderation, ii) post-moderation, iii) internal moderation and iv) external moderation.

Pre-moderation and post-moderation are directly related to the learners' assessments in terms of the qualification for which they are studying. All qualifications from the New Zealand Qualifications Framework (NZQF) consist of a certain number of credits and are designed with the inclusion of unit standards from different fields, sub-fields and domains set by the NZQA. For every unit standard that is being taught in class, learners need to be assessed against prescriptions supplied by the NZQA, which are publicly available from the NZQA website. Pre-moderation is the process of designing, checking, and/or changing assessment tasks against the outcomes stated in the relevant NZQA prescription. This stage is important to ensure that the assessment tasks are aligned with the intended learning outcomes. Research shows that misalignment between assessments and curriculum poses a threat to students' achievement (Boss, Endorf, & Duckendahl, 2001). At the pre-moderation stage, care is taken to ensure that the assessment tasks, which can be in the form of assessment booklets, assignments, tests or other forms of assessment tools, are aligned with the expected learning outcomes. Once the assessment tasks are finalised, the teaching-learning process takes place in class so that the learners can respond to the scheduled tasks. The learners' responses are usually marked and assessed by the lecturer/teacher/kiako who taught them in class. Other lecturers or teaching staff then check a small sample of these to monitor the consistency of the marking. This process is called post-moderation.

Internal and external moderation are the other two ways of categorising moderation activities, which take place within the same or different organisations. The above-mentioned processes of pre- and post-moderation, when done internally within the organisation and involving the staff of the same organisation, are termed as *internal moderation*. When the process, is conducted by personnel from a different organisation it is called *external moderation*. In some contexts, it is accepted that the teachers are likely to know more about their pupils than an external examiner, and thus are in a position to provide more information about them than a necessarily short examination can. However, teachers cannot be sure they are accurately assessing their own students in relation to other pupils in other schools, which would require either moderation or an external examination according to the Schools Council (1964, cited in Chamberlain, 1988). The rationale behind external moderation is to ensure high quality teaching-learning is achieved and maintained at the required national level, verified by a third party outside the organisation. For some TEOs in New Zealand, for which NZQA is the Standard Setting Body (SSB), this third party is the NZQA. Once external moderation is completed, the NZQA sends moderation reports to the TEOs with feedback. The reports have sections such as 'overview', 'commentary', and 'moderation results', wherein the overall feedback is summarised. If the external moderator is satisfied, the overview section states that the assessment materials meet the national standard and in the commentary it mentions that the assessor judgments about learner performance are met. This type of feedback is commonly known as a 'pass' within the organisations. When the moderation report includes any requirement of change, it is known as a 'fail'.

Although the two processes of internal and external moderation have the same purpose, the different perspectives applied to these processes by the TEOs and NZQA often create confusion and lead towards a 'fail' in external moderation. Not only that, lecturers at the TEOs often disagree with the NZQA moderation decisions (NZQA, 2010).

This study was conducted to investigate the processes of internal and external moderation and to understand the reason for these points of disagreement. The aim of the study was to identify the challenges that are faced in internal moderation and develop strategies to enable TEOs to have greater success with external moderation. The research questions involved queries in three major areas: 'What are the usual issues raised by the external moderators?', 'What are the challenges faced in internal pre- and post-moderation?' and 'How can we achieve better and more effective moderation processes?' These queries were addressed by looking at the way moderation takes place within an organisation. The qualitative research paradigm was adopted. The case study was designed by following Robert K. Yin's (1984) design. The research investigated a contemporary phenomenon

within its real-life context, when the boundaries between phenomenon and context are not clearly evident. Multiple sources of evidence were also used in this research.

One Private Training Establishment (PTE) from Auckland was selected for this case study. This PTE has five departments, one of which is Early Childhood Education (ECE). The ECE department administers the National Certificate in ECE (level 5). The background research for this paper analysed the moderation activities of this department in 2010. The Annual Moderation Plan 2010, which included nine unit standards from the qualification (see Table 1), was investigated in detail.

Table 1: Nine Unit Standards from Annual Moderation Plan 2010 at the Selected PTE

Field, sub-field and domain	US Number	Unit Standard (US) Title	Level	Credit	Status	NZQA prescription used	Version available
Education > Early Childhood Education and Care > Early Childhood: Professional Practice	9293	Describe and reflect on expectations, and develop own philosophy, of professional practice in an ECE service	6	6	-	Version 2	Version 3
Education > Early Childhood Education and Care > Early Childhood: Educational Theory and Practice	10025	Demonstrate knowledge of protection for young children from abuse	3	2	Expiring 31 December 2014	Version 3	Version 3
Sciences > Home and Life Sciences > Food Technology and Nutrition	6632	Apply knowledge of age-related nutrition needs in providing food for a child	2	5	-	Version 3	Version 3
Education > Early Childhood Education and Care > Early Childhood: Educational Theory and Practice	9321	Practise safe supervision of young children in an early childhood setting	5	3	Expiring 31 December 2014	Version 2	Version 3
Education > Early Childhood Education and Care > Early Childhood: Educational Theory and Practice	10029	Demonstrate knowledge of theories of human development across the lifespan and their relevance to ECE practice	6	10	-	Version 2	Version 3
Education > Early Childhood Education and Care > Early Childhood: Professional Practice	9297	Discuss and reflect on maintaining own health, wellbeing, cultural safety and professional integrity in an ECE service	6	5	-	Version 2	Version 3
Education > Early Childhood Education and Care > Early Childhood: Educational Theory and Practice	9301	Demonstrate knowledge of DOPs (1-5) for learning and development in a chartered EC service	5	6	Expiring 31 December 2014	Version 2	Version 3
Education > Early Childhood Education and Care > Early Childhood: Educational Theory and Practice	9324	Develop emergency plans and procedures for early childhood settings	5	4	Expiring 31 December 2014	Version 2	Version 3
Early Childhood Education and Care > Early Childhood: Family, Whānau, Community, and Society	9332	Facilitate early childhood education and care arrangements with whānau/family	6	5	-	Version 2	Version 3

Note: Table developed by the researcher using NZQA prescriptions

Content analysis and interviews were the two research tools used in this study. For content analysis, the following documents were used relating to the nine unit standards from the Annual Moderation Plan 2010:

- 9 NZQA Prescriptions
- 9 assessment booklets (old) - OASB
- 9 assessment booklets (new) - NASB
- 9 student information booklets (old) - OSIB
- 9 student information booklets (new) - NSIB
- 9 external moderation reports from previous year - EMR
- 18 moderation meeting reports - MMR

The documents listed above were the secondary data source for this research whereas the ECE teaching staff was the primary data source. The documents were used as this study looked at the moderation processes that took place in 2010 in the ECE department of the selected PTE.

In the following section the characteristics of the secondary and primary data sources are discussed briefly.

The secondary data sources comprise documents related to nine unit standards from the qualification of National Certificate in ECE and Care (level 5). For a clearer understanding of this study, it is helpful to analyse the locations of the selected unit standards in the NZQF. The sample of this research deals with two fields from the 17 identified in the NZQA framework,¹ Education and Sciences. Each field on the framework has sub-fields, for example, the Education field has eight sub-fields.² The sub-field Early Childhood Education and Care is the most significant one in this study. Again, each sub-field has domains where there are a number of unit standards for each domain – the sub-field early childhood education and care has four domains³ from which only two involve the unit standards selected in this research, Educational Theory & Practice and Professional Practice.

The characteristics of the primary data sources are summarised below (see Table 2).

Table 2: Characteristics of Primary Data Source

Participants	Gender	Ethnic group	Experience in NZ TEO	Experience in the selected PTE	Type of contract
P 1	Female	Tongan	1yr. 3mths	3 months	Full-time
P 2	Female	European	10 years	5 years	Part-time
P 3	Female	European	2.5 years	2.5 years	Casual
P 4	Female	European	7 years	1.5 years	Part-time
P 5	Female	European/English	6.5 years	6.5 years	Full-time

Note: P stands for research participants attending interviews

¹ Agriculture, Forestry and Fisheries; Arts and Crafts; Business; Community and Social Services; Computing and Information Technology; Core Generic; Education; Engineering and Technology; Health; Humanities; Law and Security; Manufacturing; Māori; Planning and Construction; Sciences; Service Sector; Social Sciences (NZQA, 2011a).

² Adult Education and Training; Adult Literacy Education; Early Childhood Education and Care; Educational Administration; Generic Education and Training; Pacific Islands Early Childhood Education; Special Education; Teacher Education (NZQA, 2011a).

³ Early Childhood: Educational Theory and Practice; Early Childhood: Family, Whānau, Community and Society; Early Childhood: Home Based Caregiver Management; Early Childhood: Professional Practice (NZQA, 2011a).

The secondary data were collected at the first stage of data collection. The NZQA prescriptions, old assessment booklets and old student information booklets were collected from the departments' academic folders. For the purpose of internal moderation and bringing changes into the old booklets, two moderation meetings were held to discuss each unit standard in 2010 starting from March and running till September. The minutes of those 18 meetings were collected from the administrative folder of the ECE department. Also the external moderation reports for all those nine unit standards from the year 2009 were collected.

An e-mail invitation was sent to the ECE teaching team asking for their voluntary participation in the study. All of the ECE teaching staff (five out of five) participated in the interviews and gave their valuable opinions.

The data collected from both the primary and secondary sources were analysed using the *general inductive approach* (Bryman & Burgess, 1994; Dey, 1993; Thomas, 2003). The documents were read thoroughly. Firstly, the external moderation reports were examined and issues were identified and noted in the first column of the content analysis form. Then the moderation meeting reports, old and new assessment booklets and student information booklets were examined to see how the issues were discussed and resolved, and this was recorded in the next two columns. The sources, for example, MMR 9/07/2009, were noted besides each finding.

The trustworthiness of findings was assessed against the feedback from participants in the research. A seminar was arranged to disseminate and discuss the research findings and four out of the five participants attended the seminar, where they analysed and acknowledged the research findings. Also the research findings were compared with, and found to be similar to, a previously conducted survey by the NZQA (2010).

The case study, which was conducted in its natural setting and reveals information in detail "recognizing its complexity and context" (Punch, 1998, p. 150), identified concerns at both the pre-moderation and post-moderation stages which were categorised as conceptual, ethical, preparatory, realistic, organisational, and structural. The table below shows an initial recording of the issues from each category.

Table 3: Issues identified in content analysis

Concern category	Tally	Frequency
Conceptual	////	5
Ethical	///	3
Preparatory	//	2
Realistic	///	3
Organisational	//	2
Structural	////	4

Conceptual issues arose from the attempt to reach consensus on interpretation of the wording in the NZQA documents. NZQA prescriptions and level descriptors are the two main documents that are used in the pre-moderation stage. The NZQA prescriptions include information on the version of the prescription, title of the unit standard, level, credit, purpose, pre-requisites/entry information, special notes or explanatory notes, outcomes, performance criteria (PC) / evidence requirement (ER), range and Accreditation and Moderation Action Plan. The content analysis of this research shows that at the pre-moderation stage outcomes and PCs/ERs are always looked at (the old assessment booklets), whereas the level, range and/or special notes/explanatory notes are sometimes overlooked or misunderstood, resulting in a failure to pass external moderation. Also, sometimes the teaching staff struggled to reach agreement over the meanings of PCs/ERs, and the level descriptors were also found confusing and lacking clear direction. Another conceptual issue arose from the discussion of selecting three students' completed tasks for moderation. Ideally, the lecturers should select learner A, B and C on

the basis of high, medium and low performance. One difficulty is interpretation as the unit standard-based certificate programme requires identifying learners as 'competent' or 'not competent', but not judging their level of performance against their achievements. Special notes or explanatory notes such as the Treaty of Waitangi,⁴ and the age-related stages of early childhood also raised some issues. Understanding the Treaty of Waitangi in the education context was also an issue. Individuals' own understanding of the Treaty varied and a lack of clear guidance as to how to incorporate the Treaty in tasks was felt with two of the unit standards, 9293 and 9332. Unit Standard 9293 is the unit standard related to developing own teaching philosophy. It was found difficult to incorporate the Treaty in the task asking for the influential factors from the learner's life that contributed to developing her/his own philosophy. Yet the Treaty was added to the task instruction to satisfy the external moderator, not knowing how international students who had been living in New Zealand for only two months could incorporate the Treaty of Waitangi in their answers. Unit standard 9332 relates to care arrangements between the ECE service and the family of the child attending. Again, not knowing clearly how the Treaty of Waitangi can be included in a task asking students to write a settling-in policy, it was included in the task instruction anyway. Similarly, it was thought difficult to include age-related stages for every PC/ER in 10025 (the unit standard related to child abuse) as the steps to seek professional help would be the same for the infants, toddlers and young children.

The following points relate to structural issues.

Lecturers were recruited as teaching-staff at the TEOs. As tertiary teachers, they certainly have teaching, administration and pastoral care responsibilities. Analysis of moderation meeting minutes (and also interviews) revealed that the participants of this study questioned these responsibilities. They were unsure under which category moderation responsibilities fall.

The structural issues can be briefly discussed as follows:

- For internal moderation, checking the same booklet every time does not take much time. That is why there is no payment for this. However, should it be part of a lecturer's responsibilities? It is the lecturer's responsibility to photocopy three samples. Is it the lecturer's responsibility to moderate others' photocopied work too? Since no payment is involved and also it is not specifically identified as a teaching responsibility, the processing of internal moderation is often irregular.
- The authority does provide guidelines. However, these contain few examples.
- There are two parts in each external moderation report. One checks the booklet (pre-moderation stage when at the institute); the other checks the photocopied sample of assessed students' work (post-moderation stage at the institute). Would it help to achieve more effective post-moderation if these tasks were separated?

Realistic issues:

- Having word counts in the Judgement; is it possible for all tasks?
- Photocopying and keeping records: sometimes it is difficult and may be unidentified, such as when assessment materials include play resources, presentations and big posters.
- Some unit standards have both theory and practical components. Sometimes it is difficult to keep track of, and photocopy, both theory and practice materials (practicum booklet) of the same student as they do them at different times.

⁴ A treaty signed in 1840 between the British Crown and the Māori in New Zealand (Orange, 1987).

Ethical issues:

- Internal moderation: should it be done only when required by NZQA external moderation?
- It is easier and more effective if the moderator has had the experience of teaching the same paper. So should the lecturers take turns to teach all papers? Yes, for moderation. No, for wisdom.
- The Treaty of Waitangi: Should it be used just for the sake of using it? In some cases, it is more tokenism than real. Is it always relevant or required? As discussed under the conceptual issues, how authentic a response could international students or even many of the domestic students produce including The Treaty of Waitangi as an influential factor on the development of their own philosophy of life?
- External moderators' comments and focus often vary. Sometimes it is felt that the TEO teaching staff are at the mercy of the moderator's own particular viewpoint.

Organisational issues:

- The checklist for internal moderation is brief and does not align with the external moderation report. Why is it not aligned?
- The newly recruited staff does not understand the process well.

Preparatory issues:

- Training or workshop to share for internal moderation
- To what extent is assistance from NZQA available?

Findings from the interviews

These issues are explained in more detail in the interviews, parts of which are included here. According to the research participants, the challenges for internal post-moderation included five factors.

Challenges of post-moderation:

- Workload and responsibilities
- Sample selection and differences in approach
- Combining in-class and practicum components
- Late submission of assessments
- Newly appointed lecturers and their understanding.

The first issue arose from the flexible, consumer-friendly setting of a small-scale PTE. Often the late submission of assessments by learners makes the marking late and thereby delays the process of moderation:

I do it regularly but I don't regularly complete the process ... the input is on time but the output could be delayed.

Often students get re-submissions from the first marking. This means lecturers need to wait until the second or third submissions to get three completed and credited samples for moderation.

This process becomes even more complicated when two lecturers share the teaching and marking of a unit standard.

The second issue is related to the selection of samples. Teachers shared different personal strategies to interpret and select from the categories of high, medium and low performance:

High is obviously the high achievers and they're generally the people who put in their assignments on time. The low is – they're the obvious – the people who are struggling, not necessarily with the content of the unit standard, but it's the English. So the medium is just finding someone in between these two.

Usually the students high and low are fairly easy because they're very obvious; and medium is usually where most of the students are, so then I just try and vary the students. I am aware of being careful, maybe not to always choose the same high person....try and vary it a little bit.

Their answers are put together well, the information is comprehensive and it's just a joy to read; and you just know that these are the kind of students that you want to represent your teaching. The other ones, I get a student that does fairly well. I try and get students who have had no resubmits because by the time you're getting into resubmits it's getting messy and they haven't really grasped the point, so I nearly always get ones with no resubmit. Then I select one that....maybe just over the kind of threshold of being acceptable, but perhaps hasn't put a lot of work in or maybe their English.

They have also mentioned selection by convenience because of a limited choice:

...because not all the students submit their work on the due date. So then you actually need to select from the bulk that you have

Right now I have only three students (in my class), and I think there is no choice.

Having unit standards with theoretical and practicum components was another challenge for the NCECEC. As the visiting lecturers may vary and the students do these at a different time, it gets difficult to ensure all practicum tasks are post-moderated alongside the classroom components.

Time spent on moderation was another important challenge mentioned by the interviewees. There are a number of actions for moderation that take the lecturer's time:

...deciding who you're going to moderate, the photocopying, getting them to the other teacher....so mainly it's time.

For newly recruited lecturers, the content and the marking can also be a challenge in the internal moderation process. The assessment booklets given to them are new to them. The time available for moderation is not enough to research the contents separately. Lecturers described different ways to overcome the challenges to resolve the problems they face.

Teachers' own strategies to manage late submissions and shared teaching of a unit standard included reminding students of the due dates in class, being patient and remembering to keep copies of moderation after checking, marking at home, liaising with the relevant teachers, and keeping a log on the computer.

The challenge of selecting student samples is resolved through personal interpretation as discussed before.

The organisation now requires teachers to keep copies of students' practicum booklets regardless of whether his/her theoretical work is selected for moderation or not. This has resolved the difficulty of accessing and combining theory and practicum components:

I think we have excellent procedures in place.

However, some suggested that there could be other ways to do it:

I would suggest that the practical and theory component can be marked together. This is when I like the students to still reflect on theory components, when they complete the practicum component. That's what I am thinking of. It would be fine to

combine the two and then mark together so that the students would reflect well on the same unit standard.

I think possibly, exemplars of practicum practice, could be included within the teaching. I think it might make it easier for them to make the transition from the classroom to the centre.

The implementation of this suggestion would be complicated, as the whole procedure would then be delayed even more. However, the ECE department in the selected PTE is still refining the process to give the best result:

What we do is; essentially we don't actually internally moderate the practicum task. We probably should. We internally moderate the class taught, part of it. The practicum tasks, the teacher who visits the student and who marks the practicum book photocopies all the practicum tasks and puts them in a filing cabinet – either in a Prac 1 or a Prac 2 sleeve. When it comes time to get the assessment books ready for external moderation then we chase up the practicum tasks to add to them, otherwise they just stay in that drawer.

Acceptance of the moderation as it 'needs to be accepted' is the usual answer to describe how the challenges related to time and workload are resolved. Lecturers more or less agreed that they accept this 'extra' work by managing their workload:

Just have to be prepared to stay longer after classes to get them in.

...there is not really much you can do about it. It just has to be done.

Well, the extra work - I just make a space and I just do it patiently and just accept it takes a long time.

A few lecturers mentioned doing the moderation as part of the teaching process, whereas others didn't see the value of moderation. Contradictory responses were received from the lecturers in the same organisation:

You just manage it. You just have to do it. It does add to our workload and nobody enjoys it; and nobody feels good about it because you don't get any extra money for doing it.

I'm the teacher – I'm the first one that is interested to see the result of my teaching and marking, so I consider it as part of my teaching.

Some participants saw the moderation process as part of their job, as they gained experience within the organisation.

I gradually became aware that it's part of my job; I just fit it in really.

It always has been part of my responsibility, not long after I started with them.

These newly recruited teachers, who raised the issue of moderating other lecturers' marking without a thorough comprehension of the process or the content, admitted having comparatively more careful attitudes towards the process of internal moderation.

Challenges in pre-moderation were more with time and experience:

More time would be needed; experience would be needed. I think, only teachers who had actually taught many times... and for a while these unit standards and they have done practicum visits as well. So combining the theory and practicum can actually be involved.

Discussion

NZQA publishes an upgraded manual every year to assist the TEOs with effective moderation, for example, the National External Moderation Manual for Tertiary Education Organisations (2011 TEO Manual). It explains the process and requirements for national external moderation of unit and achievement standards managed by the NZQA (2011c). The study conducted here reiterated those requirements. The researcher suggests a model that would provide guidance for effective moderation, in line with the findings of this research (Fig. 1).

Who	Lecturers & assessors	TEOs	NZQA
	Social moderation involving staff meeting	Identifying responsibility at staff orientation stage	Ongoing inter-active process to assist external moderation
How	Elements to consider from NZQA Prescription: <ul style="list-style-type: none"> • version • level • specifiable note or explanatory note • outcomes • evidence requirement (ER) and • range 	Allocation of time to staff to upgrade to the later versions	Availability of more exemplars
	Individual strategies for regularity	Development of annual internal moderation plan	Glossary of terms
Why	Value of moderation: Quality education and meaningful teaching-learning-assessing		

Note: Model developed by the researcher rationalised from the research findings

Figure 1: A Model for Effective Moderation

The 2011 TEO manual (NZQA, 2011c) includes a list of people for whom the guide will be useful. This includes chief executive officers, heads of departments, academic or quality managers, internal moderators, moderation liaisons, teachers/tutors and assessors. The model (see Figure 1) has identified three stakeholders where lecturers and assessors are the people directly related to the learners' learning and assessments. They often also work as internal moderators as in the selected PTE. The term TEOs may stand for people who are in management positions in the organisation, such as chief executive officers, heads of departments, and academic or quality managers. At the selected PTE, the moderation liaison person is the principal and director who also comes under the TEOs. In the model, NZQA stands for the external moderators and the material developers who are staff of NZQA and are involved in developing documents such as the TEO manual for external moderation.

The middle section of the model includes strategies that might be useful for effective moderation. Social moderation or discussing things by arranging moderation meetings with all or a majority of the staff members is a helpful strategy for effective moderation. The respondents in the survey conducted by NZQA (2010) also identified 'a collaborative approach' and 'regular *tah*⁵ meetings' as an important factor that positively impacts on the quality of moderation within the organisation. Another important 'how' strategy for lecturers when designing or changing assessment tasks is to consider every part of the NZQA prescriptions, not only the learning outcomes. Furthermore, the staff assigned as moderators need to find individual ways to track the required moderation activities

⁵ Working as one, according to Māori custom

effectively. TEOs need to help staff by providing detailed training and clarification at the orientation stage where they are introduced to the annual internal moderation plan. NZQA can provide a glossary of terms where all words are clearly explained. For example, the level descriptor states that learners at level 6 will apply a range of standard and non-standard processes. These words or phrases 'standard process' and 'non-standard process' can be interpreted in many ways. The glossary could include definition and examples of the terms used in level descriptors, the TEO manual, the NZQA website or unit standard prescriptions. Another useful strategy would be for NZQA to provide more exemplars and for TEO staff to have the opportunity to participate in face-to-face discussions. The empty sub-rows in the middle row imply that many other strategies could be included, depending on the context of the moderation.

The more significant insight from this research is given in the bottom row of the model. Often the processes and strategies get a lot of attention and the underlying value is forgotten. Moderation is all about whether the learners are being assessed correctly and fairly or not. The whole process runs around another process assessment. The term 'assessment' comes from the Latin verb 'Assidere', which means 'to sit aside' (Chapman & King, 2005). This meaning is significant in terms of the purpose of assessment. It emphasises the situation where the learner takes a solitary position to prove that the teaching-learning activities were effective for him or her to achieve certain competencies. The teachers/lecturers check how successful the learners were in their attempts. If this checking by the teachers is being checked again, which is moderation, it is important to remember that the whole process is designed not only to maintain equality and fairness, but also to improve the quality of teaching-learning. The purpose of moderation should uphold the value of quality education and meaningful learning and it needs to be shared. However, it is often ignored by all three stakeholders identified in this research: lecturers, TEOs, and NZQA.

Lecturers would do better to see the process of moderation as a useful tool that helps to reflect on their own teaching practices. It should not be perceived as a fearful event that identifies their flaws; rather it should be understood as a positive criticism of their teaching practices for future development. This approach to moderation would be useful to maintain regularity in moderation alongside the huge workload of teaching and marking. A study in Australia shows that the teachers' assessment practices, their attitudes and beliefs towards assessment impact on the way they perceive moderation (Connolly, Klenowski, & Wyatt-Smith, 2011). The study also acknowledges the values of moderation, that it supports teacher judgments, achieves fair and dependable judgments, and it is responsive to a wide range of evidence types and assessment contexts. The assessment contexts need to be dynamic. The values of quality and improvement, above compliance, need to be acknowledged by the teachers so that they perceive moderation as a means to develop their teaching and assessment practice.

TEOs need to be more specific in identifying the responsibilities surrounding moderation activities. This research reveals that often the moderation responsibilities are not clearly identified and that is how it is always put aside 'to do later'. People dealing with moderation need to know whether it is part of teaching responsibilities or administration responsibilities or a distinctive category of responsibility added to their workload. The TEOs could identify this in their quality management system. NZQA suggests any moderation system be established as part of the quality management system (NZQA, 1992). Thus moderation responsibilities should be analysed in detail on the basis of the following questions.

- Who will make the moderation policies?
- How will this be done?
- Who will implement these policies?
- How will this be done?
- Who will evaluate policies and implementation?

- How will this be done?
- What are the likely costs of setting up and operating a moderation system?
- Who will pay for this?
- How will resource effectiveness be achieved? (NZQA, 1992)

As one type of TEO, the same is to be applied to the PTEs. A PTE is defined in the Education Act 1989 as 'an establishment, other than a public tertiary education institution, that provides post-school education or vocational training' (Statistics New Zealand, 2006). The New Zealand Association of Private Education Providers (2011) has 360 members. It is the researcher's own understanding that these PTEs follow neoliberal policy that supports the marketisation of education. To make the system cost effective, PTEs may engage newly recruited staff to start, who then find that moderation responsibilities have been quietly added to their workload. The study shows that moderation responsibilities, either paid or unpaid, need to be identified clearly and at the beginning of the teacher's employment (except for those institutes that recruit people for moderation only). If moderation is not paid for, TEOs could take the approach that internal moderation is a professional development opportunity.

The NZQA is responsible for the quality assurance of non-university tertiary training providers. NZQA is the body that conducts external moderation for qualifications in New Zealand. All quality assured qualifications are set into the NZQF designed by NZQA. "Bringing coherence to New Zealand qualifications" and "introducing a fair system which measures achievement against clearly stated standards" were among the purposes stated prior to the establishment of this framework (NZQA, 1993, p. 2). NZQA should place the value of quality education above compliance issues in their actions, strategies, and guidelines concerning moderation. A survey report published in November 2010 (NZQA, 2010) states that moderation must be seen as a positive learning experience rather than an audit-type process. NZQA needs to work to establish this approach. In one of the papers at the conference for the 21st century, the credit-based system was explained during a discussion of the Assessment of Prior Learning (Slowey, 1992). It says, the credit system includes three elements which are independent of each other – a quantitative measure, a qualitative dimension, and a performance indicator. The quantitative dimension measures the amount of **credit**. The qualitative dimension measures the **level** of the learning. The performance indicator announces an individual in this system as 'competent' or 'not competent.' NZQA needs to show a value-laden approach to moderation as part of the contribution towards the individual's learning and their achievement of competencies involving both quantity and quality.

To conclude, if a shared understanding is achieved, it would be clear that although TEOs and NZQA have different functions, both are involved in the moderation process with the same purpose, which is to ensure quality education and meaningful learning for all learners.

Acknowledgements

Students and colleagues at the ICL Business School; especially to Ewen Mckenzie Bowie, Sandy Duncan, Dr. Syed Mohammad Masud Jamali, Dayal Talukder & Sarwan Kumar.

References

- Boss, T., Endorf, D., & Duckendahl, C. (2001). *Informing state assessment from the local level: A district's reflections*. Paper presented at the annual meeting of the Mid-Western Education Research Association, Chicago, Illinois.
- Bryman, A., & Burgess, R. G. (Eds.). (1994). *Analyzing qualitative data*. London, UK: Routledge.
- Chamberlain, G. S. (1988). *A reference test approach for moderating internal assessment at the upper secondary school level* (Unpublished Master's thesis), University of Canterbury, Christchurch, New Zealand. Retrieved from http://ir.canterbury.ac.nz/bitstream/10092/3497/1/Thesis_fulltext.pdf
- Chapman, C., & King, R. (2005). *Differentiated assessment strategies: One tool doesn't fit all*. Thousand Oaks, California, USA: Sage.
- Connolly, S., Klenowski, V., & Wyatt-Smith, C. (2011). Moderation and consistency of teacher judgement: Teachers' views. *British Education Research Journal*, 38(4), 593-614. Retrieved from <http://eprints.qut.edu.au/43600/>
- Dey, I. (1993). *Qualitative data analysis: A user-friendly guide for social scientists*. London, UK: Routledge.
- NZPEP. (2011, August 26). *Members*. Retrieved from <http://www.nzapep.co.nz/Members.htm>
- NZQA. (1992). *Developing a qualifications framework for New Zealand: Designing a moderation system*. Wellington, New Zealand: Author.
- NZQA. (1993, January). *A future with standards: A guide to the National Qualifications Framework and its systems* (booklet for the video). Wellington, New Zealand: Author.
- NZQA. (2010). *Summary of responses to survey of tertiary education organisations that assess against NZQA-managed standards*. Wellington, New Zealand: Author.
- NZQA. (2011a). *Explore framework*. Retrieved from <http://www.nzqa.govt.nz/framework/explore/index.do>
- NZQA. (2011b, October 13). *Assessment and moderation*. Retrieved from <http://www.nzqa.govt.nz/providers-partners/assessment-and-moderation>
- NZQA. (2011c). *National external moderation manual for tertiary education organisations: 2011 TEO manual*. Wellington, New Zealand: Author.
- Orange, C. (1987). *The Treaty of Waitangi*. Wellington, New Zealand: Allen & Unwin.
- Punch, K. F. (1998). *Introduction to social research: Quantitative and qualitative approaches*. London, UK: Sage.
- Slowey, M. (1992). Prior Learning: International developments in recognition. *International conference: Qualifications for the 21st century* (pp. 62-72). Wellington: Victoria University of Wellington.
- Statistics New Zealand. (2006). *Student loans and allowances*. Retrieved from http://www.stats.govt.nz/browse_for_stats/education_and_training/Tertiary%20education/StudentLoansandAllowances_HOTP06/Technical%20Notes.aspx
- Thomas, D. R. (2003). *A general inductive approach for qualitative data analysis*. School of Population Health, University of Auckland. Retrieved from http://www.fmhs.auckland.ac.nz/soph/centres/hrmas/_docs/Inductive2003.pdf
- TKI (2011, 11 January) *Moderation*. Retrieved from <http://assessment.tki.org.nz/Moderation>
- Yin, R. K. (1984). *Case study research: Design and methods*. Newbury Park, CA, USA: Sage.

The Assessment of Literacy Capability in Adults: Tensions and Complexities

Helen Anderson and Lisa Maurice-Takerei
Manakau Institute of Technology, Auckland

Email: helen.anderson@manukau.ac.nz; lisa.takerei@manukau.ac.nz

Abstract

This paper considers some of the tensions surrounding the assessment of adult literacy as a generic set of transferable skills. It debates the use of this kind of assessment as a performance measurement for both student achievement and quality education outcomes. Highlighted are some of the complexities associated with adult literacy measurement, the wide ranging views of what adult literacy means, and the difficulties associated with the provision of relevant, accurate, and useful measurements. The paper also reports on a small case study at an urban polytechnic and calls for more considered debate in the area of adult literacy assessment in New Zealand.

Introduction

The use of assessment information as an evidence base for teaching and learning decisions as well as a mechanism for reporting has been evident since the early 1990s (Carless, 2007; Hattie, 2003). Multiple purpose assessment tools (New Zealand Council for Educational Research (NZCER), 2006) have both a summative purpose in terms of accountability and reporting against funding requirements of the New Zealand Tertiary Education Commission (TEC), and a formative purpose in relation to enhancing teaching and learning and supporting lifelong learning goals (NZCER, 2006; O'Sullivan, Berryman, & Bishop, 2010).

The Literacy and Numeracy Assessment Tool for Adults, officially launched in New Zealand in 2010 is one such tool. The tool was developed for the TEC by the NZCER, Australian Council of Educational Research (ACER), and Fronde.

However, the use of multiple purpose assessment is not without tensions and dilemmas (NZCER, 2010). The assessment of literacy and numeracy of adults in a tertiary environment in New Zealand highlights the uneasy interrelationships of a quality education, student outcomes, and accountability evidence.

There is a wide range of opinion about what counts as evidence and how it can be used effectively in education. Indeed, the politics of what counts as evidence and how evidence is turned into data is given "amazingly ... little attention" (Denzin & Giardina, 2008, p. 16) in education. This paper explores some of the dilemmas and challenges associated with literacy assessment as an educational measurement. The paper also explores the idea of literacy testing as a useful tool to measure effective teaching and learning. More critical debate is called for in this area. The context of this discussion is the New Zealand tertiary education system where generic literacy and numeracy assessment is required in tertiary programmes at levels 1 – 3 as a condition of funding.

Creating a valid and practical assessment process to measure progress in literacy and numeracy, and hence identify the outcome of investment into literacy and numeracy development while also achieving learning and teaching value, has been a significant challenge. The tension between "proficiency" and "practice" or "generic" and "discipline" based skills adds to the complexity of the debate. Longitudinal studies have been informative in providing evidence of the different lifelong trajectories and responsiveness to intervention of generic/proficiency skills in contrast to practice/discipline skills (Reder & Bynner, 2009). The validity of generic assessment of literacy in adults is challenged by studies tracking the progress of adult literacy and the effects of various interventions (Murray, 2009).

The 'Literacy Issue'

An international measurement for skill levels in adult literacy in New Zealand is taken from the Adult Literacy and Life Skills (ALL) Survey and its predecessor International Adult Literacy Survey (IALS). The 2006 survey provides a comparative analysis of adult literacy and is undertaken across 20 OECD countries. Literacy skill profiles by the ALL survey places individual countries on an international literacy continuum and is used as a benchmark to compare New Zealand adult literacy against other comparable nations. This is a generic-based literacy tool; tasks contain "adult contexts/content" (Kirsch, 2001, p. 18) and each item is scored and scaled based on the level of difficulty. Adult literacy proficiency is placed on a scale from 1 to 5 based on responses to tasks. The tool was primarily designed to "inform policy decisions" (Kirsch, 2001, p. 9).

It has been suggested that the literacy skills of individual citizens are a powerful determinant of a country's innovative and adaptive capacity (OECD, 1997). Literacy skills are associated with individual and wider societal economic wellbeing (Walker et al., 1996) and are linked to New Zealand's "successful transition to a knowledge-based economy" (p. 6). Literacy is seen as a critical mechanism for challenging inequality through its inextricable links with physical health, access to learning, participation in democracy, and participation in the workforce (Murray, 2009).

Spurred by literacy survey results and comparisons, concerns have been raised about the ability of the New Zealand workforce to compete in a global knowledge economy. Industry, business, trade unions and providers have joined forces to outline their concerns about adult literacy in New Zealand (Industry Training Federation, 2007). In their joint 2007 publication, *The key steps forward for workforce literacy*, in their analysis, they highlight that according to IALS analysis 1 in 5 New Zealanders has poor literacy levels (Walker et al., 1996) and further suggest that the "literacy issue in the adult working population" (p. 3) is a challenge, one that requires a plan for "intensive provision" in the workplace (p. 15).

Such "literacy crisis" discourse (Benseman, 2003; Black & Yasukawa, 2011, p. 219) is not unique to New Zealand; similar debates occur in the UK, Australia, Canada and the US as governments and policy makers seek to find ways to improve the literacy proficiency of their adult and adolescent populations in readiness for workplace related literacy demands. Research in these countries has covered a number of areas relating to literacy programmes, practitioners, assessment and learners. These include responses to literacy 'embedding' in educational programmes (for example, Hodge et al., 2006; Roberts & Baynam, 2005; Smith & Gillespie, 2007), educator professional development for both literacy and vocational practitioners (Mackay, Burgoyne, Warwick, & Cipollone, 2006), 'how to' guides for practitioners (for example, McCaffrey, Merrifield, & Millican, 2007; Roberts et al., 2005), and literacy curriculum design (Lumsden, Leung, & Fritz, 2005) and analysis (Taylor, 2009).

However, there is little in the way of large-scale, high-quality research in New Zealand. Benseman (2003) suggests that "there is much scope for good quality rigorous research that serves both practitioners and policy makers" (p. 5). A large scale government policy response to the adult literacy and numeracy concerns in New Zealand is seen in the development of a literacy and numeracy assessment tool to be implemented nation-wide. This is a mass exercise which employs pre and post-testing at the start and end of all level 1-3 courses that attract funding within any programme of study within a tertiary education organisation (TEO).

The assessment tool has a model similar to the IALS. It is made up of reading tasks taken from 'adult contexts/content' and students are scored from 1-6.

The Literacy and Numeracy for Adults Assessment Tool

A key government priority in tertiary education is to improve literacy, language and numeracy skills outcomes. Improved literacy and numeracy is linked (The New Zealand Tertiary Education Strategy, 2010-2015) to quality teaching and improved completion rates. It is a government expectation that polytechnics will support students with low literacy, language and numeracy skills to improve these skills and progress to higher levels of learning. To this end, “closer monitoring” of literacy, language and numeracy improvement in students in level 1-3 qualifications is signalled (see p. 22).

Closer monitoring of literacy, language and numeracy improvement is underway in 2011 with the use of the Literacy and Numeracy Assessment Tool for Adults. Pre and post-tests, administered by TEOs, are generated by the TEO from a ‘pool’ of approximately 2,000 assessment items which are linked to a national database and connected to individual national student numbers (NSNs). The assessment tool reports are linked to the TEC Literacy and Numeracy Learning Progressions published in 2008. The goal is to ensure all courses within level 1-3 of the New Zealand Qualification Framework which include literacy and numeracy assess students using the Assessment Tool. By entering a student’s NSN, the types of assessments they have been assigned and the results of those assessments can be viewed. Student literacy and numeracy performance, based on student assessment results, can be tracked.

While the ability to track student performance in non-discipline-based literacy and/or numeracy tasks provides a unique nationwide monitoring opportunity, whether and how such information can be usefully connected to discipline-based teaching and learning goals for enhanced educational outcomes within a vocational education context requires increased consideration.

There are some key assumptions that sit beneath the ideas outlined in the Tertiary Education Strategy (2010-2015) relating to quality provision and literacy and numeracy. Firstly, there is the association of literacy in the adult world with a technical skill or a set of “information-processing competencies” (McCaffrey et al., 2007, p. 34), generic and independent of the context in which they are used (McCaffery, et al., 2007) and able to be measured. Secondly, there is the assumption that gains in literacy and numeracy measured by the assessment tool are synonymous with quality education. There is much work to be done to evaluate and align the resources available to the tertiary sector for literacy and numeracy development for students with effective measurable strategies that will show added value in the future workplaces of students and in their community engagement.

Literacy and the Adult World

Murray’s (2009) analysis of the New Zealand “Competent Learners” study (see http://www.educationcounts.govt.nz/publications/schooling/competent_children_learners for multiple reports from this longitudinal Wellington-based study) highlights the richness and complexity of data needed to begin to understand how literacy and numeracy progresses, how it interacts with a range of social and cognitive competencies, how to measure it and how to intervene. The “Learners” study is a follow-on from the Competent Children study that has tracked children from early childhood through to age 20 and ongoing. In conjunction with Statistics Canada, NZCER added test items from the other measures to the study testing. Key findings thus far include the resilience of early generic literacy and numeracy scores especially in the highest and lowest quartiles (Wylie & Hodgen, 2007). This suggests that tertiary education must be both circumspect about expected changes to generic scores in adults and willing to consider other emanations of literacy and numeracy that may be responsive to practice and discipline embedding.

Reading and writing in the adult world takes many forms. Conceptions of literacy have expanded (Hanifin, 2008) and there is continued debate about whether literacy is a set of skills or is linked to content and context. Literacy can be taken to refer to “acts of and

practices surrounding the reading and writing of disciplinary written texts” (Moje, 2007, p. 9) as well as a specific, generic skill set. Regardless of perspective or theoretical stance about the nature of literacy, whether it is a generic, skills-based set of abilities or is culturally and contextually embedded as suggested by Gee (2001), theorists seem to agree on two key issues: that connecting disciplinary knowledge to everyday knowledge is neither obvious nor straightforward, and basic reading skills do not automatically evolve into more advanced reading skills (Shannahan & Shannahan, 2008). It is suggested that specific texts from specific disciplines have specific demands which require specific practices (H. Anderson, 2011).

The notion of “literacies” as proposed by the New London Group offers a perspective built from socio-cultural theories of learning that may assist in the conceptualisation of literacy and numeracy as socially embedded skills with no authentic life outside of their everyday practice domains (Cope & Kalantzis, 2000; Pahl & Rowsell, 2005). It is suggested, therefore, that the measurement and interpretation of adult literacy proficiency is complex and contains a number of variables (Brown, Prisuta, Jacobs, & Campbell, 1996). Furthermore, the National Research and Development Council for literacy and numeracy, UK, suggest that external assessment instruments are unlikely to reflect gains made in literacy since adult literacy progress is personal and individual (Brooks, Heath, & Pollard, 2005).

Not enough debate has occurred around the use of this kind of assessment to measure literacy and/or numeracy gains in adults involved in tertiary level programmes. Moreover, if debate were to occur, dispute would be located in the beliefs educators and administrators have about how people learn, skill transfer, and the nature of adult literacy and numeracy.

Quality Education, Vocational Education and Literacy Measurement

If literacy progression is to be a measure for quality teaching and learning in level 1-3 tertiary courses as suggested in the Tertiary Education Strategy (2010-2015) it is worthwhile considering the purpose of vocational education and the nature of quality teaching and learning within the vocational context. Similarly, consideration of the best mechanism by which to measure progression requires substantial debate.

It is acknowledged that both the concept of teaching quality and the measures used to identify quality are problematic, complex and contestable (Goe, 2007; Ingvarson & Rowe, 2008). The notion of what quality teaching at this level and in this context might consist of is little understood. Often ‘quality’ teaching in this sector refers to either compulsory or university education (Robertson, 2008) and there is a general lack of recognition of the specific pedagogies that exist within vocational education (Gleeson, 2005; Maurice-Takerei & Jesson, 2010). Indeed quality in this arena is a “complex phenomenon for which no general and absolute agreement exists...” (Goe, 2007, p. 12).

While there is much literature on quality teaching and learning, there is little “empirically based knowledge about effective teaching in vocational education” (Lynch, 1997, p. 59). Recent Australian studies by Yvonne Hillier (2009) and Jane Figgis (2009) have explored innovative and effective teaching in a broad-brush manner and have focussed on what good vocational educators should do.

Figgis (2009) suggests that traditional roles and subjects in this area of education have changed and continue to change. Accurately assessing the roles and purposes of tertiary vocational education, especially at level 1-3, is not straightforward.

Furthermore, measuring teacher quality through measurable student outcomes highlights an important area of contestation in the area of teacher quality. Ingvarson and Rowe (2008) recognise that such ‘econometric’ models have led to misleading interpretations of quality teaching. Goe (2007) suggests that “Measuring teacher quality using standardized achievement test scores is challenging” (p. 8).

Data are meaningful and useful when educators 'own' the data. Educators within the vocational education sector have a strong identity within their field of expertise. Literacy data that report on and measure content and skills that do not seem immediately relevant to the trade or occupation can challenge or threaten the "occupational identity" (Seddon, 2009, p. 66) of educators as recognition of the skills and knowledge of the vocation take second place to what can be viewed as 'imposed' skills and knowledge. Whether and how educators can own the data from the TEC assessment tool remains an area of concern.

Mediation - An Alternative Approach

Millward and Timperley's (2010) study of the effectiveness of data-informed literacy professional development in New Zealand schools emphasises the need for there to be a clear connection between the data collected, their constructive analysis, and conversion into an action plan. Where generic literacy and numeracy data may be acceptable for accountability reporting, for them also to be meaningful, useful and a mechanism for change or transformation in teaching and learning, it requires mediation. If it is possible to do so, a plumbing tutor will need to translate the information into plumbing terms, a carpenter into carpentry terms, and a hairdresser into hairdressing terms that are meaningful, useful and relevant to specific educator practice. Finding ways for educators to use the data as a unit of change for relationships between educators and students and to impact on pedagogical content knowledge involves a process of meaning-making (Krauss, 2005; Lincoln & Guba, 1994).

Similarly, Shannahan and Shannahan (2008) suggest that the specialised literacy demands within specific disciplines requires the involvement of practitioners. Such practitioners, they suggest, require good initial training in an education programme that prepares them for the rigours of education at this level and which includes improved training in the literacy and numeracy of their discipline. An "archaeology of the disciplines" and disciplinary practices (Moje, 2007, p. 35) that "mines both the cultural practices and cognitive processes" (p. 36) relating to a discipline will support the development of improved literacy-based practices.

A Case to Consider

This paper was prompted initially by the experience of the authors in implementing a literacy and numeracy embedding strategy into an urban polytechnic. The impetus came from the funding availability and the awareness of the need to design approaches that would support vocational educators to build their literacy and numeracy related teaching capabilities to enhance student outcomes.

There have been three key elements to the strategy, each having had its own tensions. Collectively they interact and thus demonstrate the complexity of the task. The task is best defined as building systems to ensure continuous improvement of student outcomes with regard to effective preparation for/transition into the workforce and participation in the wider society. Adult literacy and numeracy capability is an essential embedded sub-set of this goal. The three elements of the strategy are: professional development for tutors; data gathering to inform decision making and to meet accountability requirements; and student engagement.

Professional development is a persistent theme where the organisation provided a number of opportunities for tutors to engage in professional development clusters, formal courses, and communities of practice. All tutors teaching on level one to three courses are required to participate and this has been generally well accepted and enjoyed as enhancing the experience of teaching and supporting outcomes for students. The use of testing (development of tests and interpretation/use of test data) has been a significant part of the professional development offered.

Method

This case has been designed to generate discussion from a single site, mixed-methods, authentic practice data collection across quantitative data, practice descriptions, and focus groups. Data collection has been carried out with regard to the measurement of literacy: pre and post-testing results for six courses were analysed for the Literacy and Numeracy Assessment Tool for Adults literacy results, producing a mean “step” change (TEC, 2008). Spearman Correlation comparisons (SPSS V 17) were carried out to consider the similarity/difference in the pre and post-test results. The contextualised test results were available for four of the courses and similarly analysed, and course results were noted. The purpose of these analyses was to consider whether the two forms of literacy tests were operating as proxies to measure student progress in literacy in the given contexts and to consider the students’ course results as a direct measure of student progress in the context.

Informal moderated focus groups (G. Anderson, 2004) were held immediately following assessment. The goal of the focus group was to record student experiences of each of the assessments and to thus gauge and consider student engagement. Students who had undertaken both assessment tasks were invited to discuss their impressions of the assessments and to respond to three questions. Participating students were asked what they did and did not like about each of the assessments and what comparisons they would make between the two assessments. The questions were delivered by the moderator and written on a large white board. The moderator recorded student responses on the white board as a mechanism to ensure accuracy. Each response was checked in a verbal manner with students. Qualitative data were gathered from two student focus groups: ten students from a level 2 course and 22 students from a second level 2 course. Each course had a range of ages with an even spread of mature students returning to study and school leavers.

Findings

Student Achievement Results

Table 1 demonstrates that neither the assessment tool nor the contextualised testing provided any consistent measure of change in students’ literacy. The similarity of scores suggests both test types are measuring within the same construct. These data show that these fully moderated courses had very positive pass rates.

Table 1. Pre and post-test measures of literacy in six level 1-3 programmes

Programme	N (1)	Mean step change (2)	Correlation pre and post-tests (3)	Correlation pre and post-tests (4)	Enrolled/Pass rate by courses (5)
Programme A	23	.08	.703**	.577**	91%
Programme B	13	.04	.754**	.714**	82%
Programme C	22	.03	.538**	.452*	94%
Programme D	14	0	.843**	.243	49%
Programme E	62	-.04	.388**	NA	89%
Programme F	16	.06	.464	NA	93%

Sig **.01 two tailed. Sig*.05 two tailed.

1. Number of students with complete data sets. Note that only those with complete data sets were used in the analysis.
2. Mean step change using the assessment tool measurement (TEC, 2008).
3. Correlation between students’ pre and post-test steps on the assessment tool.
4. Enrolled pass rates for students with complete data sets in each course analysed.

Student Impressions

Positive student impressions related to the usability of the assessment tool. Students suggested that the tool was “well presented” and “user friendly”. Usability aspects were identified in a positive way; “highlighting the words was good”. Students considered that the assessment was in some way good for them, for example one student stated that they thought it was “good for my IQ” and another suggested that it is “a good thing to practise.”

In response to the things they didn't like about the assessment tool students tended to talk about how they felt and what they didn't understand. They also responded to the content. There was, for example, “out of it questions that don't relate to anything” and “stuff I didn't want to read, like the government stuff.” Students suggested that there was “too much information” and identified particular readings that they didn't understand. One student suggested that the assessment was “OK for scholars.” Many of the more negative impressions related to the way students felt during the assessment; the assessment was “frustrating”, “long and boring”, “nerve-wracking”, “tiring” and also “disappointing.”

When invited to compare the assessments, students tended to focus on the readability of the assessments and the relevance of the subject matter, for example, “when [it] was related to (vocation) I understood [it]” and “[it] made more sense.”

Discussion

This case provides a set of findings limited in its generalisability by its focus on authentic data collection and modest numbers; however, the purpose has been to prompt discussion and closer scrutiny of some of the assumptions that have become part of an institute and national process.

Tutors bring their vocational identity to the task of teaching students as their primary impetus. Therefore, professional development with regard to literacy and numeracy will make sense when it is in terms of the content/skills of the courses they teach. In this part of the strategy identifying context and establishing authentic embedding has been a challenge, especially when the pre and post-testing is presented using general “adult world” content.

The approach taken of developing local contextualised tests to ensure tutors own the process and have a useful base for pedagogical and curriculum decision making has had successes and challenges. Most recently we identified that while content authenticity is vital, the literacy/numeracy skills tested must also match the cognitive demands of vocational skills and that these vary across vocations and within subsets of vocations. There is much work to be done to identify these rather than relying on unmatched generic cognitive skill sets that have not been authenticated against the work practices/cognitive demands of specific vocations.

This dilemma is deeply connected to the provision of test data that will meet accountability requirements but also inform tutor decision making. Where there is a conflict between external requirements and tutor needs, motivation is impacted. When we considered data from the “tool” testing, we found little change over the time of a course and this may be predicted from longitudinal studies that demonstrate how persistent generic test scores are both for children and adults, noting that the privileged “generic” skill set is not authenticated as relevant for the diversity of work, community or social contexts but makes assumptions about commonality and transferability. Critical theorists have long identified this process as maintaining a stratified society rather than building diverse capability.

The most important element of this work has to be student engagement. A drop in literacy achievement results between pre and post-assessments was noted in some cases. Tutors reported a number of students leaving the post assessment early or completing it in a

haphazard or *laissez faire* way. It is suggested that students who found the assessment too long, too hard or lacking in relevance, as noted above, tolerated the first round of assessment and switched off in the post-assessment round. Unless much work is done to relate the assessment to the vocational environment in which students are embedded, disengagement can result. Poor scores as a result of student disengagement were discouraging and demotivating for educators.

While some educators readily accepted the idea that comprehension or vocabulary assessed using non-discipline related reading material interrelated with the reading demands of their discipline, others were less likely to consider that the reading skills and demands assessed by the tool related to their own specific area or discipline.

The literacy and numeracy scores generated by the assessment tool student reports did provide a basis for enhanced educator reflection, opportunities for discussion, and the detection of some skill sets in students not previously noted.

Alongside the generic testing several courses implemented “embedded” testing where the tests were tutor constructed and tied to course content. Analysis of the results of these tests suggested that they were tapping into the same construct of literacy as the “tool”. After some discussion it was concluded that the questions in these tests had been guided by the formats of the “tool” and consequently were not tapping into the cognitive practices being taught and thus were not likely to improve given no overt and direct teaching intervention. The next task, therefore, is to work with experts in each vocation to build a clear and usable description of the thinking demands of their practice.

This case offers a brief look at the challenges and dilemmas of the task of assessment of adult literacy capability within its multiple contexts – political, social, vocational, pedagogical and resource driven – as seen in one TEO. No solutions are offered but the need to discuss, research and build effective practice is well evidenced.

Conclusion

This paper has explored some of the tensions involved in the measurement of existing adult literacy competency and of the outcomes of interventions. International thinking, research and practice, and the work of New Zealand researchers and practitioners has been applied to the context of the introduction of the Literacy and Numeracy Assessment Tool for Adults within tertiary level 1-3 courses in New Zealand. The New Zealand experience has mirrored many of the issues identified and discussed internationally. This paper proposes that debate in New Zealand must be robust and evidence informed to allow the practice of adult literacy and numeracy measurement for accountability and for learning to evolve positively, and that this debate must include consideration of the data and experiences found in implementation sites.

References

- Anderson, G. (2004). *Fundamentals of educational research* (2nd ed.). New York: Routledge Falmer.
- Anderson, H. (2011, April). *Insiders and outsiders: Writing for inclusion in higher education*. Paper presented to the AERA Conference, New Orleans.
- Benseman, J. (2003). *Literature review of New Zealand adult literacy research*. Auckland: Centre for Continuing Education, School of Education, University of Auckland.
- Black, S., & Yasukawa, K. (2011). A tale of two councils: Alternative discourses on the 'literacy crisis' in Australian workplaces. *International Journal of Training Research*, 9, 218-233.

- Brooks, G., Heath, K., & Pollard, A. (2005). *Assessing adult literacy and numeracy: A review of assessment instruments*. London, UK: National Research and Development Centre for Adult Literacy and Numeracy. Retrieved from http://www.nrdc.org.uk/publications_details.asp?ID=23
- Brown, H., Prisuta, R., Jacobs, B., & Campbell, A. (1996). *Literacy of older adults in America: Results from the National Adult Literacy Survey*. Washington: National Centre for Education Statistics, US Department of Education.
- Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57-66.
- Cope, B., & Kalantzis, M. (Eds.). (2000). *Multiliteracies: Literacy learning and the design of futures*. London: Routledge.
- Denzin, N., & Giardina, M. (2008). *Qualitative inquiry and the politics of evidence*. California: Left Coast Press.
- Figgis, J. (2009). *Regenerating the Australian landscape of professional VET practice: Practitioner-driven changes to teaching and learning*. Adelaide, Australia: NCVET.
- Gee, J. (2001). Reading as situated language: A sociocognitive perspective. *Journal of Adolescent & Adult Literacy*, 44(8), 174.
- Gleeson, D. (2005). Learning for a change in further education. *Journal of Vocational Education and Training*, 57(2), 239-246.
- Goe, L. (2007). *The link between teacher quality and student outcomes: A research synthesis*. Washington, DC: National Comprehensive Centre for Teacher Quality.
- Hanifin, T. (2008). 'It's more than reading and writing': The nature and extent of adults' literacy issues. In J. Benseman & A. Sutton (Eds.), *Facing the challenge: Foundation learning for adults in Aotearoa New Zealand* (pp. 125-135). Wellington, New Zealand: Dunmore.
- Hattie, J. (2003). *Teachers make a difference. What is the research evidence?* Adelaide, Australia: Australian Council for Educational Research.
- Hillier, Y. (2009). *Innovation in teaching and learning in vocational education and training: International perspectives*. Adelaide, Australia: NCVET.
- Hodge, R., Casey, H., Jupp, T., Grief, S., Ivanic, R., Lopez, D. ... McNeil, B. (2006). *You wouldn't expect a maths teacher to teach plastering: Embedding literacy, language and numeracy in post-16 vocational programmes: The impact on learning and achievement*. London: National Research and Development Centre.
- Industry Training Federation. (2007). *The key steps forward for workforce literacy* (with Business NZ, NZCTU and Workbase). Retrieved from <http://www.itf.org.nz>
- Ingvarson, R., & Rowe, K. (2008). Conceptualising and evaluating teacher quality: Substantive and methodological issues. *Australian Journal of Education*, 52(1), 5-35.
- Kirsch, I. (2001). *The International Adult Literacy Survey (IALS): Understanding what was measured: Educational Testing Service, Statistics and Research Division Research Report*. Princeton, NJ. Retrieved from http://edcounts.squid.net.nz/_data/assets/pdf_file/0005/6665/SampleTasks-IALS.pdf
- Krauss, S. E. (2005). Research paradigms and meaning making: A Primer. *The Qualitative Report*, 10(4), 758-770.
- Lincoln, Y., & Guba, E. (1994). Competing paradigms in qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-117). Thousand Oaks, CA: Sage.
- Lumsden, J., Leung, R., & Fritz, J. (2005, June). Designing a mobile transcriber application for adult literacy education: A case study. *International Association for development of Information Society (IADIS), International Conference Mobile Learning*. Qawra, Malta.
- Lynch, R. (1997). *Designing vocational and technical teacher education for the 21st century: Implications from the reform literature*. Washington, DC: Office of Educational Research and Improvement (ED).
- Mackay, S., Burgoyne, U., Warwick, D., & Cipollone, J. (2006). Current and future professional development needs of the language, literacy and numeracy workforce. (An Adult Literacy National Project Report). Adelaide, Australia: National Centre for Vocational Education Research.

- Maurice-Takerei, L., & Jesson, J. (2010). Nailing down an identity: The voices of six carpentry educators. *New Zealand Journal of Teachers Work*, 7(2), 156-170.
- McCaffrey, J., Merrifield, J., & Millican, J. (2007). *Developing adult literacy: Approaches to planning, implementing, and delivering literacy initiatives*. UK: Oxfam GB.
- Millward, P., & Timperley, H. (2010). Organisational learning facilitated by the use of student achievement data. *The International Journal of Learning*, 14(3), 239-250.
- Ministry of Education. (2011). *New Zealand Tertiary Education Strategy, 2010-2015*. Wellington, New Zealand: Ministry of Education.. Retrieved from www.minedu.govt.nz/TertiaryEducationStrategy
- Moje, E. (2007). Developing socially just subject-matter instruction: A review of the literature on disciplinary literacy teaching. *Review of Research in Education*, 31, 1-44.
- Murray, T. S. (2009). Longitudinal research. In S. Reder & J. Bynner (Eds.), *Tracking adult literacy and numeracy skills* (pp. 85-104). UK: Routledge.
- New Zealand Council for Educational Research. (2006). *Assessment for foundation learning: The importance of purposeful assessment in adult literacy, numeracy and language courses*. Retrieved from <http://www.nzcer.org.nz/research/publications/assessment-foundation-learning-importance-purposeful-assessment-adult-literacy>
- Organisation for Economic Co-operation and Development. (1997). *Literacy skills for the knowledge society: Further results from the International Adult Literacy Survey* (Technical Report). Paris: Author.
- O'Sullivan, D., Berryman, M., & Bishop, R. (2010). *Scaling up education reform: Understanding the politics of disparity*. Wellington, New Zealand: NZCER Press.
- Pahl, K., & Rowsell, J. (2005). *Literacy and education: Understanding the new literacy studies in the classroom*. London: Sage.
- Reder, S., & Bynner, J. (2009). *Tracking adult literacy and numeracy skills: Findings from longitudinal research*. UK: Routledge.
- Roberts, C., Baynham, M., Shrubshall, P., Brittan, J., Cooper, B., Gidley, N. ... Walsh, M. (2005). *Embedded teaching and learning of adult literacy, numeracy and ESOL: Seven case studies*. London: NRDC.
- Robertson, I. (2008). VET teachers' knowledge and expertise. *International Journal of Training Research*, 6(1), 1-22.
- Seddon, T. (2009). The productivity challenge in Australia: The case for professional renewal in VET teaching. *International Journal of Training Research*, 7(1), 56-76.
- Shannahan, T., & Shannahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, 78(1), 40-59.
- Smith, C., & Gillespie, M. (2007). Research on professional development and teacher change: Implications for adult basic education. In J. Comings, B. Garner, & C. Smith (Eds.), *Review of adult learning and literacy: Connecting research policy and practice* (Vol. 7, pp. 205-245). Retrieved January 7, 2008, from www.ncsall.net/fileadmin/resources/ann_rev/smith-gillespie-07.pdf
- Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21-36.
- Tertiary Education Commission. (2008). *Learning progressions for adult literacy and numeracy: Background information*. Wellington. New Zealand: Learning Media.
- Walker, M., Udy, K., Pole, N., May, S., Chamberlain, G., & Sturrock, F. (1996). *Adult literacy in New Zealand: Results from the International Adult Literacy Survey*. Wellington: Ministry of Education.
- Wylie, C., & Hodgen, E. (2007). *Competent learners @ 16: Competency performance and development over time*. Wellington: NZCER.

Formative Learning: A Description of an Assessment Framework in Initial Teacher Education

Ann Balcombe, Liz Everiss, and Margaret Brennan
Open Polytechnic

Email: abalcombe@paradise.net.nz; liz.everiss@openpolytechnic.ac.nz

Abstract

This paper describes the development and delivery of a formative assessment framework embedded within a blended delivery initial teacher education qualification. The assessment framework was developed as part of the new Bachelor of Teaching (Early Childhood Education) degree at the Open Polytechnic. The developers took a coherent approach with alignment of the assessment philosophy and methodology, the programme's principles, theoretical frameworks, and pedagogy. The assessment framework promotes formative and sustainable notions of assessment 'for' and 'as' learning supported by current assessment theory and practice. Students receive targeted feedback and become fully involved in deciding what needs to be done next and where to seek guidance. Initial formative tasks prepare students for a final major summative assessment covering all the learning outcomes in each course. Additionally, the incorporation of sustainability as a principle of the programme is based on teaching and learning that is reciprocal and collaborative in nature, where peers work together in a supported way that enables collective consideration of the question "*Where to next?*" This framework puts the theory into practice. Early analysis of students' retention and achievement results suggest that this is a sound and effective approach to assessment.

Keywords: *Formative assessment, summative assessment, formative learning, sustainable assessment, reciprocal feedback*

Introduction

The formative assessment framework described in this paper is embedded within the newly developed Open Polytechnic Bachelor of Teaching (Early Childhood Education). As a specialist national provider of open¹ distance learning, the Open Polytechnic serves a distinct constituency of adult learners who want or need an alternative to face-to-face study. The majority are in the workforce. Nearly all seek a flexible form of study that enables them to balance their learning with other commitments in their lives, and to learn in their own time and at their own pace. Students entering the Bachelor of Teaching (Early Childhood Education) fit this profile. They are typically women from diverse backgrounds and geographical locations in the 25-50 year-old age group who seek to study part time and already work in early childhood services. Two-thirds of currently enrolled students identify as Pākehā.

The Open Polytechnic has, for the last 10 years, offered a Diploma of Teaching (Early Childhood Education). A strength of this programme, which continues with the degree, is the regional support provided for students by lecturers who live and work in students' local communities nationwide. Lecturers support students locally with face-to-face practice and workshop requirements and teach distance courses nationally according to their areas of specialist knowledge.

The decision was taken at an early stage to make the Bachelor of Teaching (Early Childhood Education) programme web-enhanced and to use a blended learning approach designed for distance learners utilising online support and face-to-face teaching alongside written learning guides. It was conceptualised as a closely woven, integrated programme

¹ 'Open' used in this context means that students are able to enrol throughout the year in most courses.

with each course resting on, and linking closely with, all others in the qualification. The first intake of year one students was in February 2010. There are two annual student intakes.

As a new qualification the Open Polytechnic Bachelor of Teaching (Early Childhood Education) offered a 'blank slate' to the degree developers and an opportunity to take a 'whole' programme approach. The developers were able to rethink assessment, its fit to the programme in terms of congruency between programme principles and aims, assessment tools and processes, and to draw on current assessment thinking, particularly the value of formative assessment as a learning tool for students. Formative assessment was integrated into the design of the assessment framework to be applied to every course at each level of this three-year degree. The opportunity the development of this qualification provided to think more broadly on assessment, and to create change in assessment thinking through involving students, has in turn created an opportunity to involve them as active agents in their own assessment and learning. The formative approach in conjunction with the assessment framework builds student capacity to make judgements about their own learning and also supports their longer term learning.

Literature Review

Assessment plays a fundamental role in teaching and learning and is not simply a process of comparing performance against standards – assessment 'of' learning (New Zealand Qualification Authority, 1991). Assessment, whether written or oral, is a crucial aspect of the teaching and learning process and must be considered as part of the teaching framework (Gipps, 2002). It can be formative, summative, or a combination of both and include self and peer evaluation. Formative assessment informs learning and becomes part of the teaching process while summative assessment determines students' accumulative knowledge in relation to programme standards and requirements (Garrison & Ehringhaus, 2007; Margrain et al., 2005-06). However, of all the things that make a difference to student outcomes, the power of feedback is paramount (Hattie, 2009).

Notions of 'assessment as teaching' are supported by sociocultural learning theories which "...go beyond a conceptualisation of learning as the acquisition of 'content', and instead as a process of learning" (Crossouard, 2009, p. 11). Assessment is an integral component of student learning experiences and improvement in learning should be a principal outcome of assessment (Gipps, 1999). Furthermore, people learn best when they can assume control of their own learning. This implies that it is not just teachers who need assessment skills but also learners, as the purpose of assessment is to support learning (Absolum, Flockton, Hattie, Hipkins, & Reid, 2009). The degree programme supports sociocultural and constructivist understandings of learning. Assessment tasks are designed to encourage students to draw on their personal context and experiences and to promote active involvement in their own learning. Additionally, learning activities are aligned with learning outcomes and assessment criteria provide constructive feedback to further support students' learning. Biggs (1999) has described this approach as 'constructive alignment'. While the major part of the assessment for each course in the degree qualification is 'constructively aligned', assessment draws equally on Vygotsky's (1978) ideas of allowing the use of auxiliary tools (including adult support) and aiming at producing 'best' rather than 'typical' performance, termed 'dynamic assessment' (Gipps, Pickering, McCallum, & Hargreaves, 2006).

Black and William's (1998) work created far-reaching interest in formative assessment and argued that assessment 'for' learning is the purpose of formative assessment. Although assessment 'of' learning or 'assessment for summative purposes' remains central for employers, the public, and policy makers (Margrain et al., 2005-06), assessment 'for' learning through feedback and feed-forward has been widely adopted (Dfes, 2004; Higgins, 2004). Assessment 'for' learning supports the programme's constructivist approach through asking both lecturers and students to take responsibility for the students' learning.

Black, Harrison, Lee, Marshall, and Wiliam (2003) identified elements of practice necessary to formative assessment from their work in English schools, such as the quality of questioning and feedback, sharing criteria, self and peer assessment, and the formative use of summative tasks (Crossouard, 2009). Carless (2006, 2007) highlighted the difficulties of providing interactive formative assessment between student and teacher, and the need to make formative assessment manageable for practitioners. Formative assessment practices require more frequent teacher/learner interaction and an element of learning for teachers as they learn more about the needs of their learners (Ecclestone, 2000).

Gipps (1999) defines assessment 'for' learning as a way of supporting learning and as a collaborative interactive exercise between lecturer and student to produce the student's 'best performance' (Nuttall, 1987). Feedback from the teacher to the student becomes the link between assessment and learning (Earl & Timperley, 2008; Hattie, 2009). Assessment 'for' learning requires that process should be assessed as well as product, that assessment should be dynamic rather than static, and that attention must be paid to the social and cultural context of both learning and assessment (Gipps et al., 2006). In practice, this means that learning activities undertaken by lecturers and/or their students provide information to modify teaching (Black & Wiliam, 1998). These activities and feedback best monitor progress to inform both teaching and learning. Current ideas underpinning assessment 'for' learning are based on how students learn best and include lecturers having a clear understanding of what students are trying to learn and what is expected of them, students receiving feedback about the quality of their work and what they can do to make it better, and being fully involved in deciding next steps and who can help if they need it (Black & Wiliam, 1998; Gipps et al., 2006; Margrain et al., 2005-06).

More recently, assessment discourse has shifted to assessment 'as' learning that identifies feedback to students as being central to the teaching and learning process (Hattie & Brown, 2008). With assessment as learning, students can reflect on and monitor their progress, and in turn, this practice informs their future learning goals. Both assessment 'as' and 'for' learning enable students to understand their own learning and the goals they are aiming for through effective feedback (Elwood & Klenowski, 2002). In reality, both summative and formative elements are present in most assessment feedback (Higgins, 2004) in that feedback comments are primarily formative and ultimately relate to summative tasks or grades. Higgins reminds us of what is important about formative assessment; it must foster student learning and form part of a continuous cycle of learning. Feedback from assessment should provide to students an indication of their achievements and provide them with information and guidance from which they can improve their future learning.

Boud (2000) introduced the concept of 'assessment for sustainability' and his sustainable model incorporates constructive, reflexive learning and the development of informed judgement in learners. Sustainable assessment is: "...assessment that meets the needs of the present without compromising the ability of students to meet their own future learning needs" (Boud, 2000, p. 151). Boud's thinking emphasises equipping students for lifelong learning by providing them with guidance and support as they develop the capacity to make judgements about their own learning and engage with learning communities beyond their own.

The Open Polytechnic's Bachelor of Teaching (Early Childhood Education) assessment framework incorporates Boud's (2000) concept of sustainability as a principle and, in this respect, expects the relationship between lecturer and student to be one which is reciprocal and collaborative, where the two work together, sharing and critiquing practice, a commitment to the building of a community of learners. In addition, it provides space and structure for peers to work together in a supported way.

The Assessment Framework

The assessment framework structure is designed to promote students' deep engagement with course content through drawing on formative feedback in a way directly relevant to their learning. Carless (2006, 2007) terms this approach learning-oriented assessment by reconciling formative and summative assessment to focus "all assessment on the development of productive student learning" (p. 79) through assessment tasks as learning tasks (Biggs & Tang, 2007; Carless, 2006, 2007; Gibbs, 2006), student involvement through self and peer assessment (Boud, 2000; Carless, 2006, 2007; Sadler, 2002), and closing of feedback loops (Carless, 2006, 2007; Gibbs, 2006; Hattie & Timperley, 2007). For the feedback loop to be closed, feedback needs to be acted upon. This requires lecturers to provide students with timely and targeted feedback. The structure also produces a consistent and transparent approach that is critical in a dispersed, open distance learning environment.

Philosophical Approach

The programme developers adopted a philosophical approach informed by sociocultural understandings of teaching and learning, te ao Māori, and the concept of sustainable assessment. They took the position that formative assessment is social in nature, is always located within both the teachers' and learners' sociocultural context, and assessment must have both a 'for' and 'as' learning orientation. Such an approach sits comfortably with the concept of reciprocal and contextualised learning evident in kaupapa Māori and sociocultural theory (Pere, 1991, 1994; Vygotsky, 1978) and is also reflected in Boud's concept of sustainable assessment, wherein students develop the ability to meet their own future learning needs (Boud, 2000; Boud & Falchikov, 2006). This approach highlights the reciprocal nature of learning and importance of lecturers understanding the processes of 'student knowing' and understandings rather than relying only on knowledge transmission or assessment 'of' learning (Crossouard, Pryor, & Torrance, 2004; Gipps, 1999).

Structure

The formative-summative assessment structure (refer Figure 1) is based on three steps which create a continuous cycle of reflection, feedback, and self-review. The developers adopted and adapted Hattie's (2009) meta-cognitive questions: '*Where am I at?*', '*How am I going?*', '*Where am I going?*' and '*Where to next?*' and asked students to submit reflections based on these questions with a view to supporting sustained engagement with content and their own learning. Anecdotally, we have found that this structure and process also raises lecturer efficacy and understanding of students' learning as they engage with students' narratives. Lecturers get to know how it is for students through the self-review process. The process applies Hattie's (2009) notion of 'visible learning'; teachers see learning through the eyes of students and students are their own teachers. This reciprocal engagement is based on the belief that feedback is central to teaching and learning "...an ideal learning environment is where both (lecturers) and students seek answers to each of the questions" (Black et al., 2003; Hattie, 2009, p. 271). Our early analysis of lecturer feedback and student results suggests that feedback opportunities contribute to the development of an inquiry habit of mind in the student, essentially a personal quality based on a need to know that enhances and advances understanding (Black et al., 2003; Earl & Timperley, 2008).

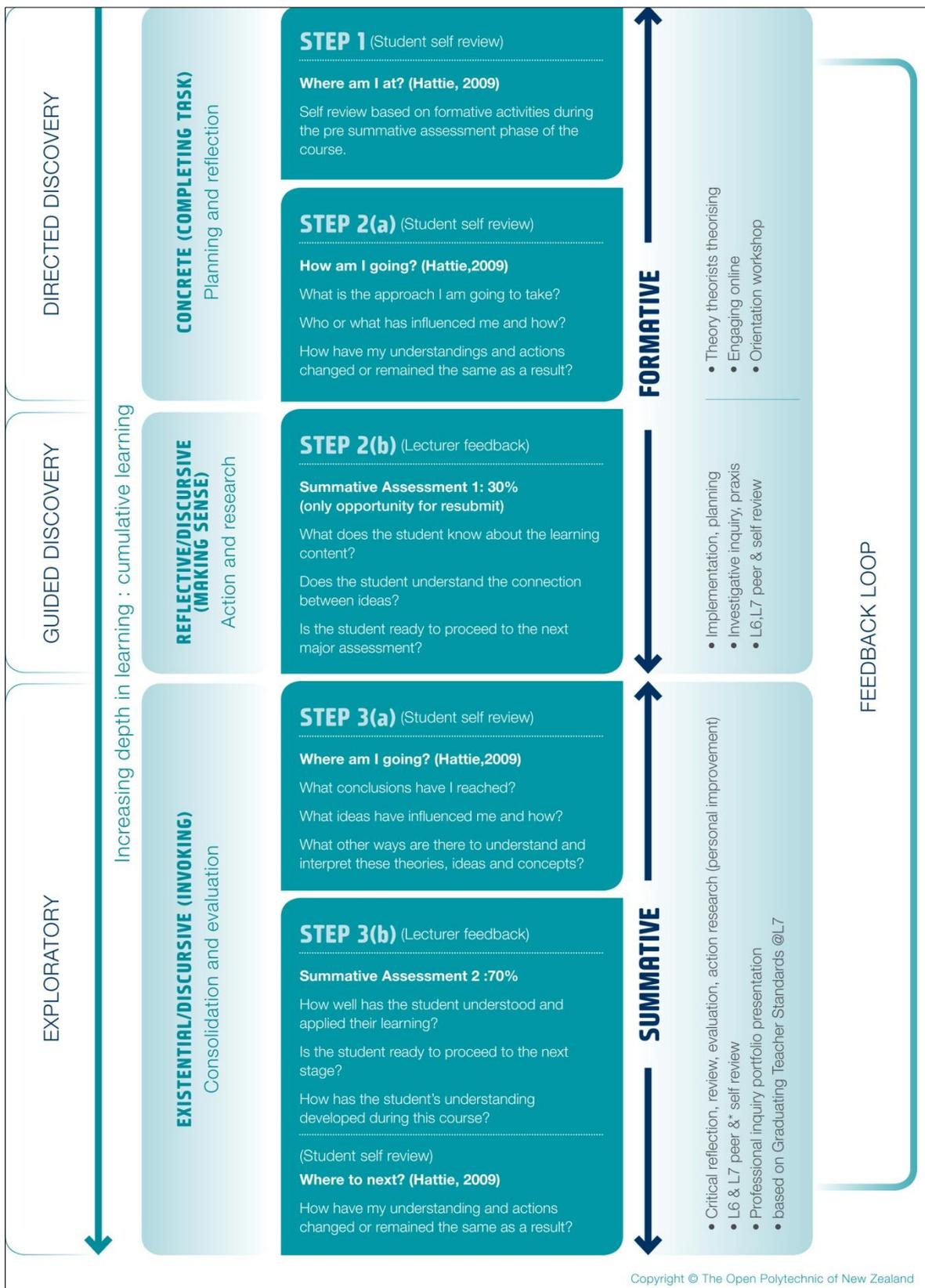
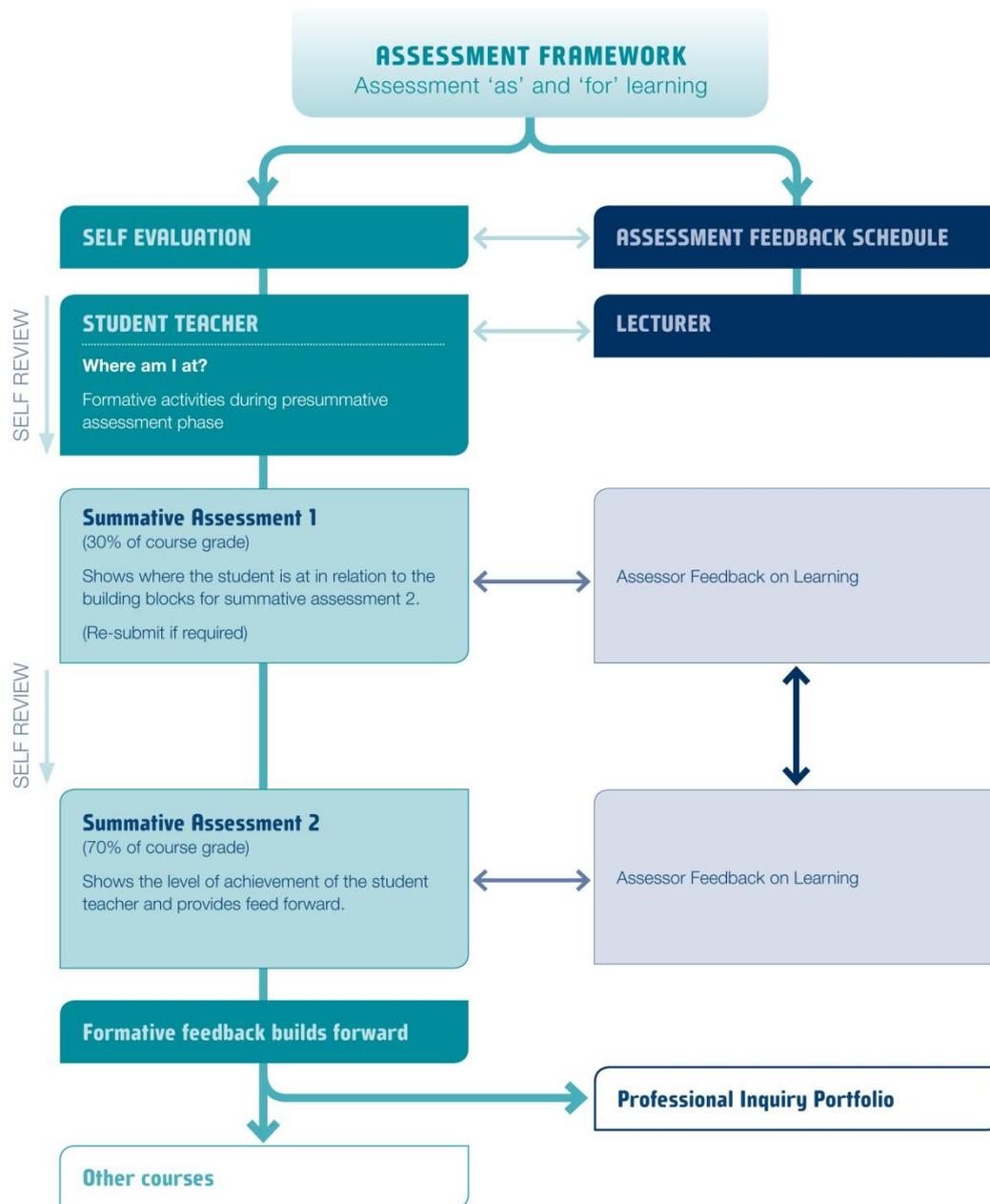


Figure 1: The Assessment Framework

Integrated Assessment with a Formative Focus

Cumulative and integrated assessment is a feature of this structure (refer to Figure 2). There are two assignments (assessments) per course with an element of iteration as the first assignment builds towards the major summative assessment in the second assignment. The formative assessment tasks are preparatory work from which the student can prepare for the next assignment task. Each assignment covers all learning outcomes, that is, learning outcomes are assessed twice in each course. Assignments involve a range of assessment events. A student may start work at any point on Assignment 1 as long as they have engaged with all the learning outcomes of a course.

Assignment 1 is worth a maximum 30% and Assignment 2 a maximum 70%. Each submission includes student self-review. Students have the same marker for both assignments. Lecturers are required to read a student's self-reviews prior to assessing their assignments and to address the self-review in their feedback to students. Lecturers provide clear instructions to students about feedback expectations and their contribution to the formative aspects of assessment. Feedback comments are primarily formative but ultimately related to summative tasks and grades (Higgins, 2004). In the degree assessment framework non-achievement in Assignment 1 still allows the student to recover in Assignment 2 and successfully complete the course.



The self evaluation is related to the student teacher's self review of their progress against assessment criteria and reflections on their learning and the planning process for the summative assessments. This is a student-directed approach to an understanding of their learning.

The purpose of submitting the self evaluation to the lecturer is to:

- Show the lecturer where the student is at;
- Provide a context for assessment feedback provided by the assessor; and
- Provide a context related to the individual student teacher to inform the assessor on the nature of the feedback required.

Copyright © Open Polytechnic of New Zealand

Figure 2: The Assessment Feedback Process

There are a number of expectations on lecturers for providing formative feedback. Lecturers are asked to:

- inform students of their availability online
- go online at least once per week to monitor, support or advise
- go online at least twice per week prior to the due dates for Assignments 1 and 2
- perform a facilitation role and an assessor role
- identify and address any teaching moments as they arise.

Resources

Resources support and clarify the formative-summative assessment approach. Rubrics are part of the assessment methodology and were developed to support this formative assessment approach and the various modes of delivery for the programme. Lecturers and students are provided with an assessment handbook containing the rubrics as guides for lecturers on how to assess, and to clarify for students how they will be assessed. The rubrics are tools with explicit criteria to guide students and lecturers to make objective, clear, consistent and defensible judgements. Whenever a lecturer or student is required to make a decision about an aspect of the assessment, there are criteria to refer to that are relevant to that performance, observation, process or product. The rubrics allow for differentiation between levels of achievement or progression and clearly describe the essence of what is being assessed and what level of quality is attached to each grade. These resources, along with the learning activities, are designed to develop important generic academic skills of self-regulation, conceptualisation, problem solving, integration of knowledge learning and multiple perspectives, progressive cognitive growth and critical reflective thinking, all of which support student success in tertiary study (New Zealand Teachers Council, 2010).

The Formative-Summative Assessment Framework in Action

The assessment framework generically applies to all courses and consists of three separate parts, each with a series of steps. Step 1 is a self-review of where the student is at, based on formative activities undertaken during the pre-assessment phase of a course. Self-review is mandatory for all students. Steps 2 and 3 involve a set of specific questions to which the student responds. The lecturer responds in Steps 2 and 3 by providing relevant and targeted feedback to guide students in their learning. The framework is used to facilitate the development of critical reflection skills and deeper learning among students.

The assessment framework draws on Hattie's conceptualisation to create a practical model. Hattie (2009) identifies the major feedback questions as being:

- Where are they going? (learning intentions/goals/success criteria)
- How are they going? (self-assessment and self-evaluation)
- Where to next? (progression/new goals)

The assessment framework adopts and adapts Hattie's focus questions as catalysts for students' thinking and writing. The process comprises three parts, each supporting students' learning and lecturers' teaching: planning and reflection; action and reflection; and consolidation and evaluation. Details of steps taken within each part of the process to facilitate the development of students' critical reflection skills and deeper learning are described in the following section.

Part 1: Planning and reflection: Where are they going?

Step one is a student self-review that involves gaining understanding of the course content in preparation for assignment writing. It is diagnostic and involves planning and reflection on what the student already knows from engaging with course content and asks these questions:

- What have I learned?
- What do I currently know about the topic?
- What other knowledge will help me deepen my understanding of the topic?
- What questions should I ask to get the information I need for myself or to offer support to others?

Part 2: Action and reflection: How are they going?

Part 2 has two components. The first (Step 2a) is a student self-review, in which the student assesses their progress and achievement level against each of the assessment criteria for a course. There is a rubric for assigning grades and another for the assessment of knowledge and skills at different levels. For an assessment the student is asked to comment in relation to their progress against the assessment criteria for the course. Assessment criteria are specific to every course in the programme.

In the first assignment students reflect on what they have learned in the course, how it has influenced them, and their approach guided by the questions: What is the approach I am going to take? Who or what has influenced me and how? How have my understanding and actions changed or remained the same as a result?

Students complete this information on a self-review form submitted with their first assignment. The form is returned to the student to support their preparation for the second major summative assessment. The student's self-review form gives the lecturer:

- an indication of the student's progress
- a context for providing assessment feedback
- an indication of the nature of the feedback required
- the student's insights about their level of understanding.

Step 2(b) is the lecturer's feedback. When providing feedback, the lecturer considers the following:

- What does the student know about the learning content?
- Does the student understand the connection between ideas?
- Is the student ready to proceed to the next major assessment?

Part 3: Consolidation and evaluation: Where to next?

Part 3 has two components. The first, Step 3(a) is students' self-review where they assess their progress and achievement level against each of the assessment criteria for a course assignment. In Step 3(a) the student reflects on what they have learned in the course and how it has influenced them. They consider:

- Why have I taken this approach?
- Who or what has influenced me and how?
- How have my understandings and actions changed or remained the same as a result of the learning I've achieved so far on the course?

Step 3(a) involves consolidation, evaluation and critical reflection. Once again students complete this information on a self-review form that is submitted along with their final major assessment. Step 3(b) consists of the lecturer's feedback. In assessing the student's level of understanding, the lecturer considers the following:

- How well has the student understood and applied their learning?
- Is the student ready to proceed to the next stage?
- How has the student's understanding developed during this course?

Pacing and Sequencing

A challenge for the programme developers was to ensure students had the opportunity to engage with course content in sufficient depth prior to assessment taking place. The content, pacing and sequencing of learning activities and assessments required careful attention and thought during the development phase. Courses are taught over a 17-week period with the first summative assessment occurring in week 10 and the second summative assessment in week 17. This timeframe allows for approximately 60% of course time to be spent on

teaching and learning through engagement with courses. Approximately 40% of the time for completing the course is specifically assessment oriented, with a focus on formative learning where peers² and lecturers provide feedback on student work (refer to Figure 1 Step 1 to Step 2(b)). The programme developers addressed this challenge by ensuring the formative aspect of assessment is embedded in all teaching and learning via feedback loops – online, face-to-face, and individualised support from lecturers. Learning outcomes are assessed twice; the first assignment builds academic skills as well as thinking; the major second assessment due to its significant weighting allows for an in-depth assessment. Lecturers and students have the opportunity to touch base with every aspect of the course and learning outcomes prior to the final submission of the major assignment.

Initial Findings and Insights

Students' high successful achievement and retention rates suggest the assessment framework produces robust assessment due in large part to the coherent and integrated approach of all aspects of the programme. Current enrolments include students who have transitioned into the degree from the Diploma of Teaching (Early Childhood Education). Year one of an initial teacher education degree programme is where students traditionally sort out whether teaching is the career for them and where tertiary providers establish the effectiveness of their programme entry benchmarks. For these reasons it is expected that completion and retention results may be lower at year one than for subsequent programme years. These results are similar to those for the Diploma of Teaching (Early Childhood Education) which also takes a 'blended' approach. However, they are more positive in that students are engaging with increased complexity in their degree studies yet are maintaining sound completion and retention rates which exceed those traditionally found in distance education (mean successful completion rate of 85% across courses). As expected, completion and retention results increase as students progress through the programme. Early retention and completion results suggest that the mean successful completion rates for Māori (83%) and Pasifika (84%) students are equivalent to the mean score for all enrolled students in the programme.

Table 1: Student Course Completion and Retention Rates

Courses	Trimester 2, 2010			Trimester 1, 2011		
	Name	No	Retention %	Successful Completion %	No	Retention %
R & I1	130	78%	75%	104	80%	79%
S in S1	79	90%	86%	96	91%	91%
L DC1	124	85%	78%	56	77%	75%
T W 1	Not offered in trimester 2, 2010			182	77%	75%
C1	170	86%	78%	64	80%	91%
R & I 2	Not offered in trimester 2, 2010			103	89%	82%
LDC2				62	90%	85%
SinS2				78	95%	91%
C3				18	94%	94%
	547	78%	78%	832	77%	82%

Key to Course Titles: R & I – Research & Inquiry; S in S – Self in Society; LDC – Learning, Development & Culture; TW – Tangata Whenua; C – Curriculum. The numeral after each title refers to the year.

² Informal peer review occurs during face-to-face group sessions: workshops, teaching practica, and online forums.

There was immediate engagement with the framework among the lecturers; they assisted with the roll-out and were motivated to make it work. The developers of the degree consulted with teaching staff and asked for feedback throughout the development process. The lecturers could see the potential of this assessment approach to foster deeper and more intense engagement with learning by students based on their prior experience. The framework encourages lecturers to ask questions relating to how to give feedback that produces a positive impact on student learning, how to build students who are assessment capable, and what influences achievement among students. It is being used as a practical inquiry model by the teaching staff. There is an awareness of the potential in distance learning for students to begin with assessment tasks and to 'find the answers' rather than engaging deeply with programme content. However, in assessment of degree courses there is a need to do all the course work to successfully complete the assessment. The consistent structure of the assessment framework across all courses has facilitated shared understandings and clarity of the assessment process among lecturers. The assessment framework structure promotes deep engagement with course content. Students appear to draw on formative feedback in a way directly relevant to their learning, and lecturers' efficacy is raised as they engage authentically with the student narrative threading through the self-review process.

Excerpts from Students' Self-Reviews

The following self-review example reflects the experience of a year 2 student across two assessments as she transitions from the Diploma of Teaching (Early Childhood Education) to the degree programme in a course which considers political, historical and cultural contexts of early childhood education and curriculum. The self-review is typical of the type of formative conversation that occurs between students and lecturers and which allows lecturers to see the impact of their teaching and feedback on students.

Year 2 Student Assignment 1 Self-Review

Why have I taken this approach? (Provide a rationale for the way you approached the assignment)

This trimester I am transitioning to the degree from the diploma and I have found the course structure and modules require a different approach from the diploma. The main difference is working through all the modules first and then working on the assignment. This is different but a more logical way of working. Discovering the 'whole' picture first I felt helped me work more methodically through the assignment requirements. Initially opening the folder and discovering all the readings left me feeling overwhelmed with the task ahead. (Year 2 Student self review, T1, 2011)

Who or what has influenced me?

Initially I found the assignment instructions hard to decipher but using the webpage and forums was very helpful to detect the key aspects required. Explanations provided by J were very helpful guidance to complete the assignment. Thank you. The set texts required for this paper were all interesting reading that I hadn't yet discovered with the diploma. In particular set text 'Facing the Big Questions in Teaching: Purpose, Power and Learning' was very influential, particularly chapter eight written by Rata, discussing inclusive and exclusive biculturalism. (Year 2 Student self review, T1, 2011)

How have my understanding and actions changed or remained the same as a result of the learning I have achieved so far on the course?

All the readings and the set text have increased my understanding about biculturalism and the exact history of the development of Te Whaariki. This paper introduced and reinforced new terminology (especially discourse), definitely increasing my understanding and knowledge through hours and hours of reading and new knowledge. (Year 2 Student self review, T1, 2011)

Excerpt from lecturer response:

Thank you ... for your thoughtful feedback. It is a 'big' course in that it covers complex, diverse yet connected topics. The readings are provided as a resource with signposts provided within the course to show which readings connect to certain content areas. You don't need to read everything in depth – skim for key points and go deeper where necessary. Because everything builds on each other you will have all the research done for your assignment when you have completed the activities. You have done well. (Lecturer feedback, T1, 2011)

Year 2 Student Self-Review Assignment 2

What conclusions have I reached?

Political influences have the biggest impact on early childhood education (ECE). The government in power dictates whether ECE flourishes with positive policies or is marginalised, with policies changed to have disadvantageous effect. A profit- driven objective is often at the detriment of advancement and goals of ECE. (Year 2 student self-review, T1, 2011)

Who or what has influenced me and how?

The work of Helen May and Anne Meade is inspirational, with the significant time and energy put into improving rights and recognition for ECE in Aotearoa NZ. My own working class upbringing in a small low economic town provides me with my own political ideas and agenda favouring a supportive state approach. Relating to writing my first ever briefing paper the feedback from L for assignment one was useful and I have tried to implement the guidance into assignment 2. (Year 2 student self-review, T1, 2011)

What other ways are there to understand and interpret these theories, ideas and concepts?

Removing myself from my own political agenda/ bias to critically analyse different ideas, approaches and concepts. Considering different viewpoints such as business/ non-business or economic versus non-economic. Treasury /business disposition of user pays, accountability, funding cuts versus the non-business viewpoints...I believe the diversity of ECE services empowers parents/whānau to choose a compatible setting matching their needs and beliefs. (Year 2 student self-review, T1, 2011)

How have my understandings and actions changed or remained the same as a result of the learning I have achieved so far on this course?

My values and beliefs about my own political agenda have been confirmed with the readings, if anything I'm even more staunch red, confirming who I will be voting for in the election. Initially I found the history factor of this course daunting; however, I really enjoyed finding out about the exact history of ECE in more detail. And feel a sense of achievement on completion of my briefing paper and I am looking forward to SinS3 next trimester. (Year 2 student self review, T1, 2011).

Hattie (2009) describes formative feedback of this nature between students and lecturers as powerful because it provides lecturers with an indication about "...what students know, what they understand, where they make errors, where they have misconceptions, when they are not engaged" (p. 7). In this instance the self-review feedback identifies key influences on the student's learning and locates the student as being in transition between programmes. This contextual information provides a useful guide for the lecturer when determining the nature of the feedback and support to provide for this student and possible gaps. It illustrates also how formative feedback and student reflection embedded in a self- review process can support and inform students on a current task whilst also having a transformative effect where students develop the ability to self-regulate immediate and future performance (Carless, 2009).

The following excerpt is taken from a year 3 student's self-review. The student's reflections indicate emerging aspirations for not only her own future learning but also her children's and provides a good example of both transformative and sustainable education.

Year 3 Student Assignment 1 Self-Review

Why have I taken this approach? (Provide a rationale for the way you approach the assignment.)

I found this assignment very enjoyable, and this surprised me! The set text by Mutch was very readable, and the course work was very supportive and clear. The workshop provided me with confidence to proceed! I had only one major concern when approaching this assignment and that was to find articles within the set list, as I found this rather limiting after exhaustive searches. I eventually sought clarification that this list was to be used as a 'guide' – so very quickly was able to find articles which were from scholarly peer-reviewed journals (and I researched the authors' credentials well before going ahead) and were very recent. The articles I have selected for this assignment are content-wise very different, but I approached this assignment with the understanding that it wasn't the 'content' that we were discussing, but the suitability of the methodology, and to demonstrate our understanding of research methods. (Year 3 Student self-review, T2, 2011)

Who or what has influenced me and how?

This assignment has given me the courage to read scholarly articles, and not be daunted by the terminology and jargon. I feel very optimistic about assignment two for this course, and look forward to working through the exercises (with fresh eyes) that we were given at the workshop by M. I feel in putting this assignment together I have done THE MOST valuable learning; my confidence is light years ahead of where I ever imagined it would be when I first set eyes on the textbooks and the word "research". After engaging with all the course material, I can honestly say that the learning has all come together for me in this assignment, and I am certain your feedback will only set me straighter and more firmly on this path. (Year 3 Student self-review, T2, 2011)

How have my understandings and actions changed or remained the same as a result of the learning I've achieved so far on the course?

Again I will state that I am very proud of myself! For the amazing leaps and bounds in my understandings of research: the differing methodologies, and all the aspects that make up and differentiate each approach, as well as the fact that the different type of method works for gathering different types of information. It has really opened my eyes, and I love "spouting out" all my new terminology! I can see the admiration in my sons' eyes "wow - mum really IS doing a DEGREE – listen to her talking like that!" (my lads are 11, 12, and 13 years old – I hope I am instilling aspirations for higher tertiary study in them!!) (Year 3 Student self-review, T2, 2011)

This student's review supports and confirms our aspirations for students based on Boud's (2000) concept of 'assessment for sustainability'. The excerpt provides evidence of the student engaging in constructive, reflexive learning but perhaps most importantly it provides an example of assessment that equips students for lifelong learning through providing guidance to make informed judgements and building confidence to engage in future learning. The student's renewed enjoyment and excitement about learning suggests that learning is most effective when students are supported well to meet their own learning goals and aspirations (Absolum et al., 2009; Gipps et al., 2006). Support can come in various forms. The programme developers' careful attention to balancing tools and strategies with interpersonal elements has encouraged the aspirational and independent learning evident in this reflection. The formative assessment focus is 'constructively aligned', yet draws equally on Vygotsky's (1978) psychological and concrete tools to support dynamic and transformative learning (Biggs, 1999; Crossouard, 2009; Gipps et al., 2006).

Discussion and Reflections on Early Analysis

The initial data from the Bachelor of Teaching (Early Childhood Education) formative assessment framework suggest that effective student learning is occurring in the programme, that students are being encouraged to assume control of their own learning, and to have 'voice' in assessment through the self-review process which is guided by the questions '*Where am I going?*', '*How am I going?*' and '*Where to next?*' (Hattie, 2009). This strong focus on formative assessment also acts to mitigate any potential conflict between practice, understanding and real experience through allowing questioning directed at assisting students rather than testing them, discussion that promotes the students' reflection on the task and its content, and on-going dialogue between the student and lecturer where the student initiates questions as well as responds (Black et al., 2003; Open Polytechnic of New Zealand, 2011).

The degree designers developed an initial teacher education qualification from inception which allowed for an integrated and coherent framework. Careful attention was paid to ensuring congruence and alignment across the programme's conceptual framework and aims, assessment methodologies, and the learning material. We see the rare opportunity to take a holistic programme approach as contributing to the success of the assessment framework. Another important contributor was the development of carefully crafted learning activities and assessment tasks. These auxiliary tools developed students' academic and cognitive skills alongside comprehension of course content (Gipps et al., 2006). Students are also given opportunities to practise skills, assess understanding, and receive feedback, prior to submission of summative assessments.

Early indications of the effectiveness of the formative-summative approach included lecturers' commitment to working with the framework. Anecdotal feedback from lecturers suggests that this allowed for a pedagogically sound way of working with students that supported independent learning in a way that felt 'comfortable' and 'right' (Crossouard, 2009). Lecturers' initial positive response has now been confirmed by students' very positive achievement results.

Emerging analysis of lecturer feedback and student results indicates that feedback opportunities contribute to the development of an inquiring disposition and a desire to continue learning (Earl & Timperley, 2008). We have found the two-way conversations where students and lecturers have shared input into the assessment process to promote a focus on learning that supports transformative learning. Put simply, the lecturer gets to know how it is for the student and is able to respond to where the student 'is at' in a way that means something for the student, rather than focusing on where the student 'should be'.

We conclude that an assessment framework that allows lecturers to respectfully respond to student learning is especially important in initial teacher education programmes because students work directly with young children and families throughout their programme of study. In addition, the New Zealand Teachers Council (2010) asks that teacher education programmes' conceptual frameworks guide all aspects of development, delivery and assessment. In our case our assessment philosophy and practice is congruent with sociocultural principles expressed in the programme's conceptual framework. This is evidenced by carefully designed learning activities that draw on students' experiences along with an expectation that students actively participate in their own learning. Student success rates to date suggest the assessment framework in conjunction with the programme approach is supporting students to engage in sustainable and meaningful learning.

References

- Absolum, M., Flockton, L., Hattie, J., Hipkins, R., & Reid, L. (2009). *Directions for assessment in New Zealand*. Wellington: Ministry of Education.
- Biggs, J. (1999). *Aligning teaching and assessment to curriculum objectives*. Imaginative Curriculum Project, UK: Learning and Teaching Support Network Generic Centre.
- Biggs, J., & Tang, C. (2007). *Teaching for quality learning at university* (3rd ed.). Maidenhead, UK: McGraw Hill & Open University Press.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Buckingham: Open University Press.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-75.
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education* 22(2), 151-167.
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long term learning. *Assessment and Evaluation in Higher Education*, 31(4), 399-414.
- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31(2), 219-233.
- Carless, D. (2007). Learning-oriented assessment: Conceptual basis and practical implications. *Innovations in Education and Teaching International*, 44(1), 57-66.
- Carless, D. (2009, November). Tertiary education in New Zealand. *Education Review*.
- Crossouard, B. (2009). A socio-cultural reflection on collaborative challenges in the States of Jersey. *Research Papers in Education*, 24(1), 77-93.
- Crossouard, B., Pryor, J., & Torrance, H. (2004). *Creating an alternative assessment regime with online formative assessment: Developing a researcher identity*. European Conference on Educational Research, Crete.
- Dfes, R. (2004). *Assessment for learning: Whole-school and subject training materials*. Retrieved December 13, 2008, from www.standards.dfes.gov.uk/secondary/keystage3/all/respub/afl.ws.
- Earl, L., & Timperley, H. (2008). (Eds.). Understanding how evidence and learning conversations work. In *Professional learning conversations, professional learning and development in schools and higher education* (pp. 1-12). UK: Springer.
- Ecclestone, K. (2000). Assessment and critical autonomy in post-compulsory education in the UK. *Journal of Education and Work*, 13(2), 141-162.
- Elwood, J., & Klenowski, V. (2002). Creating communities of shared practice: The challenges of assessment use in learning and teaching. *Assessment & Evaluation in Higher Education*, 27(3), 243-256.
- Garrison, C., & Ehringhaus, M. (2007). *Formative and summative assessments in the classroom*. Retrieved from www.amle.org/Publications/WebExclusive/Assessment/tabid/1120/Default.aspx
- Gibbs, G. (2006). How assessment frames student learning. In C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 23-36). London: Routledge.
- Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education*, 24, 355-392.
- Gipps, C. (2002). Socio-cultural perspectives on assessment. In G. Wells & G. Claxton (Eds.), *Learning for life in the 21st century* (pp. 73-83). Malden, MA: Blackwell.
- Gipps, C., Pickering, A., McCallum, B., & Hargreaves, E. (2006). "Teacher Assessment" to "Assessment Learning": The impact of assessment policy on teachers' assessment practice. In R. Webb (Ed.), *Changing teaching and learning in the primary school* (pp. 47-60). Maidenhead, UK: Open University Press.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.
- Hattie, J. (2009). The black box of tertiary assessment: An impending revolution. In L. Meyer (Ed.), *Tertiary assessment and higher education student outcomes: Policy, practice and research*. Wellington: Ako Aotearoa.
- Hattie, J., & Brown, G. (2008, September). *Assessment and evaluation*. Presentation at Speaking of Quality: An Educational Psychology Forum, Auckland, New Zealand.

- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 8-112.
- Higgins, R. (2004). *The meaning and impact of assessment feedback for students in higher education* (Unpublished PhD thesis). Sheffield Hallam University, UK. Retrieved from <http://campus.openpolytechnic.ac.nz/moodle/course/view.php?id=37570>
- Margrain, V., Everiss, L., Murphy, T., McClew, J., Edlin, A., & Meade, A. (2005-06). *Effective teacher education practice: The impact of written assessment feedback for distance learners*. Wellington: NZCER.
- New Zealand Qualifications Authority. (1991). *Standards-based assessment in general education*. Wellington: Author.
- New Zealand Teachers Council. (2010). *Approval, review and monitoring processes and requirements for initial teacher education programmes*. Retrieved from www.teacherscouncil.govt.nz/te/itefinal.stm
- Nuttall, D. (1987). The validity of assessments. *European Journal of Psychology of Education*, 11(2), 108-118.
- Open Polytechnic of New Zealand. (2011). *Lecturer assessment handbook*. Wellington: Author.
- Pere, R. R. (1991). *Te Wheke: A celebration of infinite wisdom*. Gisborne: Ao Ako Global Learning NZ.
- Pere, R. R. (1994). *Ako: Concepts and learning in the Māori tradition*. Hamilton: University of Waikato.
- Sadler, D. (2002). Ah! So that's 'quality'. In P. Schwartz & G. Webb (Eds.), *Assessment case studies, experience and practice from higher education* (pp. 130-136). London: Kogan Paul.
- Vygotsky, L. (1978). *Mind in society. The development of higher psychological processes*. Cambridge: Harvard University Press.

The Future of Secondary School Examinations: The Use of Technology in the New Zealand Context

Stephen Bargh
New Zealand Qualifications Authority

Email: steve.bargh@nzqa.govt.nz

Abstract

There is a growing international trend in the use of technology to manage external assessment from item/question development through to marking, and processing and reporting of results. It is recognised that some of the most important innovations and developments in assessment are a direct result of the technology that is available today.

External assessment for the purpose of secondary school qualifications in New Zealand currently relies on paper-based marking and a range of manual operations. This paper draws on international experience and research evidence to show how technology can improve and enhance the validity, effectiveness, and efficiency of external assessment. The paper also develops an end-to-end digital model of external assessment that could readily be implemented in New Zealand.

The model takes a two-phased approach. Phase one is the replication of the current assessment systems and processes in a digital environment. Phase two is the implementation of a range of digital tools to improve the quality of assessment.

Underlying these areas is the notion of moving from paper-based responses, to using a computer in examinations, to online sitting. It is acknowledged that the increasing use of technology in external assessment is inevitable but change of this magnitude needs to be carefully managed and planned. This paper presents a model that can be used as the basis for planned change.

Introduction

Innovation, creative use of resources, highly trained professional teachers and a wide and balanced curriculum are features of the New Zealand education system. These elements, combined with a revolutionary secondary school assessment system, have all helped place New Zealand's education system amongst the best in the world. The increasing use of technology in schools has opened up new opportunities for student learning and innovative teaching practice. But how does New Zealand compare with other countries in the use of technology in external assessment?

Scheuermann and Björnsson (2009) reported that many European countries are currently managing the transition from paper-based testing to computer based. They also note that there does not appear to be one defined pathway and that most countries are finding the transition a challenge. There is no clear solution for what is a very complex exercise.

Norway has been working through the development of computer-based testing for Norwegian reading and writing, mathematics and English since 2002. Moe (2009) related the part the University of Bergen played in these developments. They originally developed computer-adapted tests but in 2007 changed to linear testing. By 2008 the testing had been administered four times and there do not appear to be any major issues.

In the United States, while the major college admission tests, the SAT and the ACT, still rely on paper-based tests, Bennett (2011) reports that in 2010 72% of state schools in the United States reported providing online assessment. In Oregon alone 1.5 million tests are administered online each year at primary and secondary level. There are continuously available computer adapted tests in reading, mathematics, social studies and science. The state has embraced online assessment to the extent that paper-based testing is the exception.

Bridgeman (2009) reported that other large-scale testing such as the Armed Services Vocational Aptitude Battery, the Measures of Academic Progress (a school-based battery of assessments in reading, mathematics and language usage), the US Medical Licensing Examinations and the Microsoft Certification Examinations, amongst others, all make extensive use of computer-based testing.

The United States is investing heavily in the use of technology as a catalyst for changing the way they assess their students. Two major state consortia, the Partnership for Assessment of Readiness for College and Careers and the SMARTER Balanced Assessment Consortium representing the majority of state education agencies, have begun the process of reviewing their assessment processes. They intend to move to assessments that integrate closely with teaching and learning in the classroom, that have more meaning for students, and improve validity and fairness. Technology will have a major part to play in this transformation through the use of computer adaptive assessment.

In an open letter, published in a range of media in June 2011, a number of the most influential educators and education experts in the United States stated:

If done correctly, the shift from pencil-and-paper to online assessments will build upon this opportunity to transform the nation's education system and provide a platform for new approaches to learning and schooling, not just to testing. (www.innosightinstitute.org para 4)

The United States is not the first country to see the opportunity that increased use of technology in assessment would bring to their education system. Near the end of last century the Netherlands believed that the use of computers in assessment would encourage and accelerate the take-up of technology in schools. The Dutch policy makers qualified this ambition with the proviso that the assessments had to be "cutting edge and innovative and thus revealing the full potential of computer use and multimedia in education" (Hermans, 2009, p. 3). They then set about a two-pronged approach to invest in technology in schools and in high-quality assessments using the technology. CITO, the Dutch National Institute for Educational Measurement, has been heavily involved in this approach and is constantly improving and enhancing their systems. Peter Hermans (2009) of CITO reports that 90% of new assessment products are either fully or partly dependent upon computers. He also concludes that:

IT lies at the heart of the key competencies in test development and is therefore of strategic importance to any enterprise that takes the future of testing and assessment seriously. (p. 10)

In 2009, Denmark decided to take the process of using computers for assessment a step further. They allowed students complete access to the Internet for their examinations. While they are still evaluating the merits of this decision, it did provide assessment developers with the opportunity to move away from traditional examination questions and their focus on recall of knowledge. The emphasis now is on a student's ability to search for, sift through, and analyse information.

Michael Wesch, an anthropologist at Kansas State University, recently released a viral video in which he and others in the United States raised questions about the current educational paradigm. They raised issues around knowledge creation, and concepts such as the social and collaborative knowledge production were discussed. Wikipedia is an example of socially constructed knowledge, and New Zealand students need to be a part of this knowledge revolution.

These ideas change the whole focus of examinations. The current focus in New Zealand is on the recall of knowledge or the application of knowledge. The problem is that knowledge is changing at such a prodigious rate. On the Michael Wesch video Ray Kurzweil described IT capacity as increasing exponentially, and Tim O'Reilly maintains the Internet has no top to it.

Examination questions will therefore require a much different focus if they are to continue to have meaning and reflect teaching and learning. In order to realise the vision of the New Zealand Curriculum students will need to learn how to discern and filter knowledge, make connections across bodies of knowledge, and create knowledge. These processes will need to be taught in class, and external assessment will need to be aligned with these objectives. Access to the internet will therefore be crucial for learning and for assessment.

Kimber and Wyatt-Smith (2009) wrote that:

The current synergy of thinking between business, education and research suggests that today's students require a different, more complex skill set than in the past, and that their teachers have particular responsibilities in elevating seemingly superficial levels of online activity to more critical, creative, empathetic and ethical activity. Just as we can no longer think of knowledge as a fixed entity, we must find ways to carry forward those capabilities that can adapt to, critique and create newer notions of co-created knowledge. (p. 12)

Assessment will need to reflect the context that Kimber and Wyatt-Smith describe. Co-constructed knowledge or socially developed knowledge are notions that are contrary to traditional pedagogical practice where knowledge is approached as a fixed entity. Within the paradigm Kimber and Wyatt-Smith describe, assessment of knowledge becomes redundant and new paradigms for assessment must be found.

The conundrum facing examination bodies is where the push for this comes from. Ideally, it should come from teaching practice, with the assessment reflecting what is happening in the classroom. However, sometimes the assessment can provide inertia for this sort of change to occur as was the case in the Netherlands.

Computer technology has already permeated New Zealand society to the extent that it is arguably now indispensable in the way individuals work, play and communicate. The tension this integration creates is clearly seen in the education sector as schools and teachers struggle to maintain their perspective of a quality teaching and learning process and environment in the face of massive technological change.

There is already a large amount of research into the pedagogical changes required by schools and teachers to ensure their students maximise the learning potential of the technology. What is absent is how these students will cope with an external assessment that is still based on the older, more traditional pedagogy. At the very least, students who have spent three years learning in a digital environment are likely to be at a distinct disadvantage when they have to sit and answer examination questions using a pen and paper.

The retiring head of Ofqual, Isabel Nisbet, called for a change away from the traditional pen and paper examinations. She described pen and paper examinations as "invalid" for digitally native students (Stewart, 2011). Prensky (2001) refers to those who have grown up with the technology as "digital natives". They can *parallel process multitask* and see the technology as a "friend".

Bennett (2002) argues that the inexorable use of technology will lead to significant changes to the management and content of assessment. He maintains that as technology is becoming pivotal to schooling, and is the learning medium of choice for most students, using another medium for assessment is indefensible.

Brooks (2009) reported that a significant number of non-traditional learners were encouraged to return to learning through online assessments that engaged the learner through a range of interactive skills within a digitally rich environment.

There is still, however, an attitudinal divide within the education sector. Stephen Atherton from Apple spoke at the 2009 International Association for Educational Assessment (IAEA) Conference in Brisbane. He maintained that there is a disconnect between teachers and

ICT and its use. He argues that students see technology as an extension of themselves, but limitations are placed on the creative use of ICT by school and teacher policies and practices.

Atherton's view that 21st century students see technology as an extension of themselves is a critical perspective. A core philosophy underpinning the examinations in the New Zealand context is to provide students with the best opportunity for them to demonstrate a selection of their knowledge, skills and understanding. If students are more at home with technology, and are in a better position to demonstrate their knowledge and understanding through its use, then it would be an injustice to New Zealand students to continue with pen and paper examinations.

Manafy and Gautschi (2011) described digital natives as preferring to receive information quickly from multiple sources and from pictures, sounds, and video before text. They preferred to interact in real-time, use hyperlinks, and undertake learning that is instant, relevant, and fun. Digital natives are also comfortable living their lives in a public space.

Manafy and Gautschi also described digital immigrants (those who are drawn to digital technology) as preferring a controlled release of information from limited sources and they preferred this information released linearly, sequentially and logically. Digital immigrants often get their information from text and favoured a situation where they had more personal and private space for introspection.

The issue is that the processes and policies within a school are developed for digital natives by digital immigrants.

Fielden and Malcolm (2007) found that limited use was being made of cell phones for formal school activities. One New Zealand school appeared to have embraced the use of cell phone technology because it offered another form of communication and augmented other technology within the school. Fielden and Malcolm did find that where a negative policy was rigorously applied, the digital divide between the school and its students was increasing.

Some New Zealand schools have recognised the need for their students to have some form of portable technology. The recent media debate about a school north of Auckland which has made iPads compulsory is a good example of the issues schools face. The debate ranged across such concerns as the usefulness of the technology to the cost of it. The provision of iPads for students is also a good example of the ways some schools are trying to provide their students with the best opportunities to connect to the wider world.

As the world in which we all live becomes more dependent upon technology for everyday work, play and communication, and it becomes indispensable in education, serious consideration needs to be given to its use in external assessment. The question should not be if New Zealand uses technology for examinations, but how.

These items of research provide the direction that assessment in New Zealand will need to travel in order to maintain its place as an innovative educational leader. The increasing use of technology in external assessment is not only desirable, it is critical to New Zealand students for whom the digital environment is one in which they instinctively interact. Not moving to a digital external assessment system is not justifiable in the 21st century.

What is being explored in this paper is not new, as many jurisdictions already use a variety of technological tools for secondary school assessment. What is new is that this paper presents it all together as an end-to-end model for the New Zealand context.

New Zealand needs to develop a digital examination mode because it will be of benefit to the students, the digital natives who prefer to work in a digital environment. The assessment will be able to be more closely aligned to teaching and learning, thereby through examinations that are meaningful and timely. Improvements to the validity of the assessment will be able to be implemented and assessment can become more than recall of knowledge.

The Model

In implementing an end-to-end model there are two distinct phases. **Phase one** takes the current process and re-constructs it within a digital environment. This means that examination questions in a digital paper will resemble current questions, students will answer them by typing rather than by writing, and they will be marked by markers who will view the examination paper online rather than a physical script. **Phase two** will involve using the technology to expand the opportunities for students, enhance the external assessment processes, and create efficiencies for the examining body. Elements of phase two would include such things as interactive questions, adaptive testing, online open-book testing, and non-human marking.

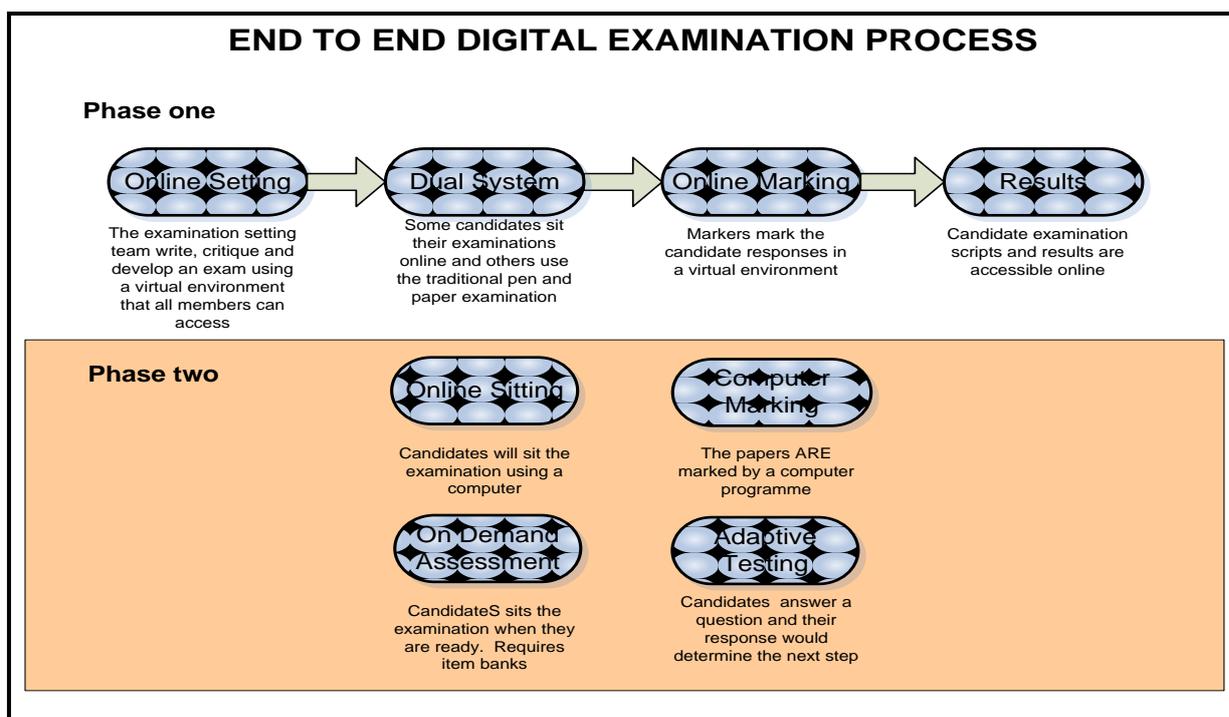


Figure 1: Diagram of an End-to-End Digital Examination Process and the Phases for Implementation

One significant concern with developing such a model is that it can only be based on known technology. A recent article in *NetGuide* (July, 2011) challenged the viability of the laptop in the future and raised the possibility that it would be replaced by tablet technology. *NetGuide* quotes Craig Richardson, the Managing Director of Jade Software, as saying “I would not say the laptop is dead, but it is definitely an endangered species.” (p. 20). The pace of change in computer technology is rapid and, while a digital assessment model will need to be based on known technology, it will also need to be flexible enough to manage technological evolution; or in the case of the tablet, technological revolution.

In moving to a fully digital examination environment the context in which it would be implemented needs careful consideration. The New Zealand community places a high value on examinations and sees them as a critical component of the educational environment. The merits or otherwise of society’s attitude towards external examinations is not the focus of this paper and can be debated elsewhere. It is sufficient to state that the examination process is highly visible and attracts significant media, political and public scrutiny.

The second contextual element requiring consideration is New Zealanders’ attitudes towards technology. While, in general terms, New Zealanders tend to embrace new technology they appear to have an inherent suspicion of it. Incidents such as the multiple

breakdowns of Telecom's much vaunted XT Network in 2010, continuing issues with Sony's Playstation Network, and other technological malfunctions have done little to assuage this general feeling of mistrust.

Any process that creates changes to an area of New Zealand educational life that is so highly valued, through the use of an element that the community is not totally trusting of, has the potential to be problematic. A careful, staged implementation plan accompanied by a comprehensive communication plan is critical to the process.

Background

New Zealand's National Qualification is the National Certificate of Educational Achievement (NCEA) which is undertaken at three levels. Assessment is standards based with a mixture of internal and external assessment. External assessment is undertaken at the end of a calendar year and the most common mode is a written examination or portfolio submission.

For written examinations a separate paper is developed on a standard by standard basis with, from 2013, no more than three standards being assessed at any one time. A setting team for each level is contracted to develop an examination for the external standards. Most members of each setting team are full-time experienced teachers. The papers undergo a series of development and critique stages until they are ready for publication. At this point they are printed and then securely stored.

The papers are sent to the examination centres prior to the examination. Most examination centres are secondary schools and they have dedicated areas where the candidates sit the examination. The candidates sit the examination and write their answers in the answer book provided. The answer booklets are collected at the end of the examination and sent to a sorting hub for distribution to their allocated markers.

Marking panels are established on a standard by standard basis and there is a series of stages that must occur before the commencement of marking. Two senior markers will conduct a benchmarking process (sometimes referred to as standardisation) whereby they will mark a selection of answer booklets, confirm the assessment schedule, and prepare for the training of markers. The panel will then meet so that markers can be trained by senior markers on how to use the assessment schedule and how to annotate. A final check of the assessment schedule is undertaken after the markers have test marked a number of scripts. Markers can then begin marking, which they do by annotating the papers. Results are written on the front of the booklet and entered online.

Booklets are collected at the end of marking and distributed back to the candidates. Results are published on each individual candidate's online Record of Achievement. Candidates have the right to appeal their results and request either a check of the marking or a total remark.

This model is focused on written examinations but portfolios cannot be discounted and most of the model can be used for the assessment of submitted portfolios. If a digital image can be created, then in most cases, the material can be electronically assessed. Computer marking would have to be evaluated on a subject by subject basis. Assessment of subjects such as digital and visual communication, and visual arts, contains an aesthetic component that software is unlikely to be able to replicate at this stage. Digital assessment of portfolios should occur when and if possible, with the aim of implementing a full assessment process at a later date.

Online Setting

Online Setting is the production of an examination paper within a virtual environment. In simple terms this means that those responsible for the setting and development of the paper produce a digital copy that is then critiqued, checked and edited within the virtual environment, negating the necessity of producing a hard copy.

The setting team do not need to physically meet but through the use of a shared workspace or a “cloud” any one of them can access the draft examination and undertake their tasks. Critiquers and checkers can also have the same access, and any reports that are required throughout the process can be uploaded in the same way. Discussions can be held using Skype or similar technology.

This process is underway and is used effectively in a large number of subjects. Anecdotal evidence indicates that most of the setting team are very happy with the process.

This process is more secure, efficient and cost effective than the process of developing a series of physical drafts. The recent earthquakes in Christchurch highlighted the benefits of using the shared workspace. Examination papers that had been developed using the shared workspace were safe.

There are several jurisdictions that use this process to produce a digital copy of the examination paper which is then printed for candidates to write in during the examination. This process, however, should be seen as an interim stage only, and the objective should be to remove the need to print a hard copy at all.

In phase one the setting team will replicate the current process in a virtual environment.

The CITO systems have evolved to the extent that their online testing tools include interactive test items and adaptive testing modules and these are in widespread use in the Netherlands. Interactive test items could include short video, movable graphs, three dimensional rotational diagrams and satellite imagery with a zoom function and the ability to overlay a range of analysis toolsets. Current examinations in New Zealand contain static content and this will remain the case in phase one, but in phase two the setting team can begin to introduce more interactive and dynamic content.

Some of these innovations can be novel and exciting. As New Zealand moves towards a more digital assessment process, care needs to be taken that the focus continues to be on the needs of the student, not on the technology. Validity is critical; the examination must test what it intends to.

Online Sitting

Online Sitting (OLS) is perhaps the most contentious aspect of the digital assessment model. It raises notions of equity of access and resourcing in education that are fraught with issues in the New Zealand education sector. An in-depth exploration of these issues is not the intention of this paper, but they cannot be ignored, and they underpin and influence elements of the model.

Decisions will need to be made in the near future regarding student access to computer technology and their use in external assessment. A number of secondary schools in New Zealand are phasing in the use of personal computer technology as a required item. Many schools are moving towards a paperless learning environment where learning is undertaken using a range of technology tools. Learners have access to the internet and all the knowledge found and developed there.

Phase one would involve the establishment of a dual system whereby some candidates would sit the examination in the traditional manner using paper and pen, and others would use a computer. In the case of the computer, the candidate responses would be typed

and submitted digitally. In both scenarios the examination questions would be set out in the same manner.

There is a range of technical issues that would need to be resolved before a dual process can be implemented. These would include security and authentication of candidate responses, technology breakdown contingencies and software compatibility. Provision of the hardware and technical support are also potentially contentious issues requiring resolution.

In some countries students are provided with a computer to sit the examination. This would involve significant expense but has the benefits of ensuring that all students have access, and reducing security issues.

The alternative is to allow students to use their own computers. Fluck, Pullen, and Harper (2009) reported in their case study of a computer-based examination how they had devised the examination using open source software loaded onto a bootable CD. Students had to load the CD in a specific way and the software locked access to the hard drive of the computer and any networking device such as WiFi or Bluetooth. Students answered the examination questions and saved their answers to the CD. The CDs were collected at the end of the examination.

One innovative idea that Fluck et al. (2009) put into action was a practice CD for the students to take home prior to the examination. They could then practise loading it onto their computer which meant they were comfortable with its use when they started their examination.

The students in the case study reported a number of issues. Several were distracted by the sound of typing and others were distracted by the need to manipulate several windows on the screen. Other issues involve appropriate methods for the collection of the student responses, and preventing students from seeing other screens. However, the case study provided a method that could be used in phase one of the implementation of the digital model, a method that is secure, prevents collusion, and allows the use of a student's own computer.

Fluck et al. (2009) found that the preference for computer-based assessment was strongly linked to prior experience in the use of computers in assessment. This aligns with Nisbet's (2011) comments about digital natives. Davies and Gralton (2009), in a trial about automated essay scoring, found that students were able to write more online than using pen and paper. They also found that the majority of students preferred to type their work rather than write it.

A process that assesses the elements of critical, creative, empathetic and ethical activity will be a part of phase two but significant testing will need to be undertaken during phase one to ensure the questions are appropriate. Online open book testing as practised in Denmark is a likely part of phase two but decisions about this should be deferred until in-depth analysis of the Danish situation is completed and decisions can be made about whether this form of assessment is appropriate for the New Zealand context.

Questions have to be asked about New Zealand society and its readiness to accept these ideas. Assessment cannot operate in a vacuum, and the social context in which these changes can be implemented needs to be carefully considered.

Online Marking

Online Marking (OLM) can be implemented with little if any dependency upon the other elements of the model. All it requires is a digital image of the student response, which can be typed text or a scanned image of a hand-written examination paper.

All usual marking processes can be utilised using an OLM programme. The marker would view the script or section of the script on a computer screen, make their judgement and input the result into the OLM programme. The result would be exported into the relevant database for reporting purposes.

It is a feature of assessment in New Zealand that the marked scripts are returned to the candidates for one last quality check. Under OLM the candidate would gain access to a digital image of their script.

OLM provides enhanced security for the scripts. Transportation of physical scripts entails a risk of loss which can create problems for the awarding body and, more importantly, for the candidate. Any reduction in the physical movement of scripts would significantly reduce the risk of loss.

In New Zealand external assessment is conducted at the end of the year and markers are under pressure to deliver the results in a timely fashion for schools and universities. Using OLM increases the time to mark for markers as the marking processes can begin the day after the examination is sat. Currently, some marking panels are unable to commence marking for up to seven days after the examination. There is also no time lag while scripts are sent for quality assurance. This can happen online and in real-time. Senior markers are able to access all markers' scripts at any time and check mark a selection. They can also undertake targeted check marking. An example of this is when a marker appears to be awarding a disproportionate number of Excellence grades. The senior marker could target this marker's Excellence results only and check mark those.

One of the initial stages of marking in New Zealand is a process called benchmarking. In this process the senior markers will select a range of typical scripts across the range of grades for a particular standard. They are also required to mark a random selection of scripts. These scripts are marked using the draft assessment schedule and this process can be used to uncover flaws in the schedule or in the paper. Using OLM would enable Item Response Theory (IRT) to be used in real-time to inform the senior markers as to how well the examination questions function and how well they discriminate between the grades. This would enable the senior markers to make informed judgements regarding the assessment schedule and, if required, any adjustments to ensure valid judgements are made.

Reporting of results under OLM can be on an item by item basis. This means that classroom teachers would be able to analyse their results on an item by item basis enabling more meaningful self-review.

There is no doubt that marking of extended handwritten text on a computer screen is different to marking a physical script. Markers report that the physical script provides them with the ability to iteratively mark, that is, turn pages back and forth in order to develop an holistic perspective of the candidate response. Online marking is generally linear as the marker moves down the screen and then onto the next page. There is little if any alternate backwards and forwards movement through the writing.

Dillon (1994) found that unwieldy onscreen navigation can reduce the marker's ability to build a sense of meaning. Remembering (and then finding) relevant factors in the student response proved more difficult using a computer screen than a physical script.

It is suggested that some of the earlier findings by some researchers were influenced by the technology available at the time. Screen resolution, refresh rates and navigation tools have improved markedly over the last decade. Johnson, Nadas, Bell, and Green (2009) found that the mode did not present a systematic influence on marker reliability. They found that markers still marked the same whether they marked using a computer or a physical script regardless of their marking profile. Some of the products that NZQA has viewed do provide for quick and easy movement between pages.

Crisp (2008) reported that despite the same training different markers focused on different elements within an extended piece of writing. She reported that, despite this, different marking styles seemed to be equally effective. Greatorex (2004) found there was no clear difference in the electronic moderation of portfolios between on-screen markers and off-screen markers. This is reinforced by David Coniam (2009) who not only found little difference, but also found that experienced markers are capable of moving to OLM with little, if any, change in marking ability.

While these studies indicate little difference between using a computer or using a pen and paper to mark, anecdotal evidence indicates a perception that there is a difference. It would be important, in the New Zealand context, to thoroughly compare the two modes within a test environment. Marker confidence in moving to a different marking system is critical to the success of a digital examination system.

In NSW, online marking is undertaken in marking centres where banks of computers are set up for markers to use. This has the benefit of centralising technical support and maintaining security. It is an expensive model to use as the resources are only used for a short period of time and it requires significant logistical support. A marking centre model restricts the selection of markers to the immediate vicinity. This operates effectively in NSW where there is a large urban area to draw from, but could be problematic in a New Zealand context.

New Zealand currently uses a domestic marking model which requires markers to use their own premises for marking. In an OLM environment markers would be expected to provide their own computer technology and ensure they have an appropriate broadband connection. Technical support becomes problematic and restricted to long distance support. This can be overcome with appropriate training and using an OLM application that is user-friendly.

One of the benefits of OLM is the ability to develop virtual panels of markers from around the country (or from around the world). All marking activities could be conducted in a virtual environment and would enable the best markers to be used regardless of their location.

Implementation of an OLM system is the key element in the digital examination process. Once in place many of the other elements can fit around it. It should be the first process implemented in phase one as it is the least threatening to the sector and requires minimal change by the student and teacher.

On Demand Assessment

One issue with external assessment, as it is currently managed in New Zealand, is the separation of the assessment from teaching and learning. In addition, feedback from the external assessment can take up to three months to be provided to the candidate. On Demand Assessment (ODA) combined with Automated Essay Scoring (AES) has the potential to eliminate this issue by enabling the candidate to sit an external assessment when they are ready and to receive feedback in a timely fashion.

ODA is used in this paper to refer to external assessment that is performed when the candidate is ready. When a candidate has completed a unit of study that contains knowledge and /or skills that can be assessed using an externally assessed achievement standard, the candidate sits an online examination that is developed and marked externally. This also means that external assessment can be used as a diagnostic tool for teachers to determine a student's strengths and weaknesses.

Item banks would be a critical factor in ODA. It should be possible for an entire group to sit an examination for a particular standard at the same time but answer different questions or different tasks. In order to be able to meet this requirement item banks will need to be developed that any examination should be able to draw from. Comparability between items is important but IRT can play an important part in the development of valid items.

IRT provides a model-based relationship between item responses and the latent characteristics assessed in a test. A range of mathematical and psychological models can be applied to evaluate whether a test item does actually do what it intends to do. IRT would be a critical component in determining the comparability of tasks developed for an item bank. Baker (2001) describes a method that he refers to as “test equating” in which a number of tests can be compared in terms of difficulty. Test equating can also be used to compare a candidate’s ability. It is this method of IRT that would be useful in terms of ensuring that items in a specific item bank are of equal difficulty.

This model readily lends itself to adaptive testing. Adaptive testing requires a response at a particular level. Most computer adaptive testing systems use IRT to draw a series of items based on a student’s ability. An item bank is essential to this process. This form of testing has the potential for diagnosing areas where students are having difficulties specific to them.

Whether adaptive testing can be used in a large-scale standards-based summative assessment process such as in use in New Zealand is a significant question. Currently, most examination items are based on extended pieces of writing and are marked holistically against the four criteria in a standard. There has been a phase-out of items that asked specific questions at a specific level in favour of more holistic items.

A decision to move to adaptive testing needs to be made on the basis that this method contributes to the improvement in the quality of assessment. In other words, adaptive testing is shown to improve the validity, effectiveness and efficiency of the assessment and, most importantly, it is a better method for the student.

To keep up with demand these item banks will require a regular developmental process and cycle. A full-time examiner and markers would be needed, especially in high demand standards or subjects. Providing them with real-time IRT tools, they will be able to improve each item using appropriate data.

He and Tymms (2005) list a range of software systems that can be used to develop item banks. While many of them are expensive, several current systems can create items banks, undertake the test, and analyse the results.

Dermo (2009) highlighted an issue that would require care. He reported that the random way in which questions from an item bank were used was considered unfair by students. Students felt that they were not being measured fairly if they had to answer different questions. The concept of comparability is critical to the development of item banks and their usage. Each item will need to be comparable to each other and a range of tools and processes will need to be introduced to ensure comparability.

The two main benefits of ODA are found in the relationship of assessment to teaching and learning. ODA allows assessment to occur when the student is ready and not at a predetermined time. Assessment can occur closer to learning which provides the candidate with the best opportunity to demonstrate their knowledge and understanding.

A possible model for ODA working in a New Zealand context would involve the candidate logging in, accessing an examination for a particular standard, responding to the items, and then submitting the examination for marking. Appropriate authentication processes would need to be developed to ensure the candidate is who they say they are.

A second benefit would be the demise of the end-of-year examinations which is a major logistical exercise requiring significant resources for an intensive period. Removing the end-of-year examinations could potentially invoke a negative reaction from the New Zealand public who appear to have a conservative perspective regarding the place of secondary school examinations in the lives of adolescents. The effect of such a removal would lead to schools changing their own cycles.

Combine ODA with AES and it would be possible for the candidate to sit the examination, submit their answers, and receive feedback and a result within 24 hours.

Automated Essay Scoring or Computer Testing

Computers have been used to mark tests for many decades. Multi-choice questions and short answer questions are the most obvious ways in which a computer could be used effectively to score a candidate and there are many jurisdictions that use these methods in their assessments. In New Zealand most external assessment makes use of extended writing tasks or paragraph type responses, both of which can be assessed using AES software.

AES makes use of a complex algorithm that assesses the extended writing against a pre-determined formula and structure. Most secondary school students approach an extended piece of writing using a similar formula which can be programmed into the computer. The computer is also trained to look for trends, flow of ideas, links between ideas, key words, and key phrases. Most of the AES programmes in use also have a fail-safe process that allows a piece of writing that might not fit the usual pattern to be assessed by a human.

Kelly (2001) reported that computer marking has several benefits. There are significant cost savings for the examination agency. Using AES with some human support is significantly cheaper than using human markers only. In New Zealand, just over 2000 markers are contracted to mark the external assessment for NCEA and New Zealand Scholarship. The costs of this exercise are considerable.

It is well known that human markers are prone to marker drift, and most jurisdictions employ some means of quality assurance to keep markers in line. A computer eliminates marker drift and is able to mark with a high degree of reliability. Kelly describes it as “virtually incorruptible consistency” (p. 4).

Davies and Gralton (2009) reported another benefit is that AES provides immediate feedback and scores. Immediate feedback to students is a critical factor in learning. Davies and Gralton undertook research into the feasibility of using computer scoring in a large-scale testing context and in the process compared human marking with computer marking.

Their study found that there was no significant difference between AES and the human markers. Their results also indicated that students preferred to type rather than write and were able to type more than they wrote by hand. Interestingly, they surveyed the students who were involved in the study and found the students were comfortable with the computer marking their work.

AES has shown to be reliable and efficient over a long period of time. Student work would need to be presented to it using typed text. This can occur through online sitting using either a keyboard or handwriting recognition tool, or through transcription of the handwritten answers. AES would be implemented in phase two.

Results

Each student’s result is currently uploaded to a database by the marker. This is a manual process requiring the marker to transpose the result from the front of the answer booklet to an online marking page. To mitigate against error, the marker is expected to enter the result twice, with the second entry being a blind entry.

Once OLM is implemented this stage will no longer be necessary as the result will be exported to the database as soon as the marker is satisfied the result is final. This process alone will save the markers significant administrative time.

OLM allows reporting of the result of each question. This means that the student will see their results on a question by question basis. In addition to the student a classroom teacher will be able to view their class results on a question by question basis, thereby providing useful feedback to them about their students and about their own teaching.

Implementation

Care needs to be taken with the implementation of the digital examination process. Phase one should be aimed at changing the current system to a digital environment with minimal effect on students and schools. Phase two will be about using the digital tools available to enhance the assessment process through ODA, interactive assessment and, possibly, adaptive assessment.

Many setting teams are already using a shared workspace for the development of their examination papers. This process has received positive feedback and will continue to grow.

Online marking is central to the entire digital examination process and should be the first element implemented. With examinations being developed digitally, the paper would only be required to be in hardcopy for the actual examination process itself. An OLM application would require the paper to be scanned and uploaded.

Implementation of an OLM process would be less threatening to the community and would require minimal change for candidates. They could still sit their examinations in the traditional manner using a pen and paper, or using a computer.

The next step would be to implement a procedure for the use of computers in examinations. This is likely to require a dual system for a period of time until it can be assured that every student has similar access to the technology.

At this point ODA with item banks can be implemented. AES can be implemented at any stage and while it could be piloted and tested using small samples, implementing it on a large scale basis for the end-of year examinations may pose risk that is considered unacceptable. AES would be better deployed in small ways throughout a period after ODA has been implemented.

Conclusion

Robert Kozma (2009), in his call to action regarding assessment of 21st century skills, listed a number of advantages brought about by the introduction of ICT into large-scale assessment. These include a reduction in the logistical costs associated with large-scale paper-based examinations, an enhanced ability to efficiently collect quality data, and to take advantage of a range of tools that are now integral to teaching and learning.

When compared to paper-based testing, use of computer technology in assessment has been recognised as providing a more valid way of assessing the skills that are required in the 21st century. Kozma also wrote that:

Traditional assessments also fail to measure all the skills that are believed to be enabled and acquired by the regular use of new, technology-based learning environments. (p. 17)

Bridgeman (2009), Bennett (2011), van Lent (2009) and Hermans (2009) have all reported similar findings to Kozma. Computer-based testing enables more successful and valid assessment of 21st century skills and competencies than the traditional paper-based assessment. If New Zealand is to fully realise the ideals of the New Zealand Curriculum in ensuring that its students are prepared for the future, then it requires valid assessments of these skills and competencies. Paper-based assessments lack the validity required.

A digital examination process will look fundamentally different to the current external assessment process used in New Zealand. Examinations may not necessarily be developed in their entirety but be a combination of a series of discrete items from an item bank. These items will have been developed by a setting team that may not have physically met but have used a digital environment to write, critique, and test the items. This process will be enhanced through the use of real-time IRT analysis to improve the quality of the items.

Appropriate software will be used to develop interactive examinations, mark examinations, and provide feedback and results. Results will be available faster and therefore feedback to students will be closer to the learning. External assessment can then be used as assessment for learning.

The implementation of technology in schools also raises questions about pedagogical and curriculum change. It is not acceptable to place 21st century technology in schools and use it in a 19th century teaching and learning, and assessment, model. The opportunities for significant positive change cannot be underestimated and these should be driven by the needs of the students, the digital natives, and not by the assessment. Care also needs to be taken to ensure that the preferences of the digital natives takes precedence over those of the digital immigrants.

Under this model the end of year examinations will cease, thereby removing the need for schools to constrain their programmes to a calendar year. Potentially, this action alone could revolutionise the way schools are structured and could lead to more personalised learning.

Removal of the end-of-year examinations will remove the need for the hugely expensive and intensive examination exercise conducted each year.

The rationale for moving to a digital model has to be underpinned by solid principles. As with the Netherlands, decisions must be made with the focus on the students, not on the technology. Changes should be made only when they are steps towards providing an enhancement of the current process. Digital examinations provide an opportunity for a range of enhancements that would benefit the student. A wider range of questions can be used that assess a wider range of abilities and provide students with more opportunities to demonstrate what they know and can do.

Note must be made that the current generation are digital natives who are used to using technology. Research noted earlier indicates that a digital environment is a better environment for the digital natives to undertake external assessment. They can type faster than they write, they prefer a digital environment, and they are intuitive users of technology. The world in which they live, and will live, is one in which extensive use is made of technology for work, play and communication. Technology permeates their lives, is an integral aspect of the way they live, and is seen as an extension of themselves. As schools make more use of technology within their teaching programmes it will soon become unjustifiable to not implement a digital examination system.

Many countries have experienced issues such as the initial high cost of implementation, unreliable technology, the range of technology used and inequality of access. These are not insurmountable but require considerable thought, planning and testing in order to be mitigated.

The context in which the digital examination process will be implemented needs to be carefully considered. Examinations are considered high stakes and have a particular status in New Zealand life. Every implementation step will require communication, not just with the sector, but with the public as well so that everyone concerned is assured that students will not be disadvantaged. Rigorous testing and piloting will have to be a feature of any implementation plan.

The New Zealand Qualifications Authority has set itself a target of achieving this within ten years. This is very achievable and should prove to be positive for students and cost effective in the long term. It is fundamental that the NZQA take a proactive stance on these developments rather than wait for a critical mass to develop before they react. The issues raised in the development of the model are not insurmountable, but will require good communication, commitment and “buy-in” from all parties concerned.

References

- Atherton, S. (2009, September). *Students and IT: A disconnect*. Keynote speech presented at the International Association for Educational Assessment Conference: Assessment for a Creative World, Brisbane, Australia.
- Baker, F. (2001). *The basics of item response theory*. Retrieved from <http://info.worldbank.org/etools/docs/library/117765/Item%20Response%20Theory%20-%20F%20Baker.pdf>
- Bennett, R. R. (2002). Inexorable and inevitable: The continuing story of technology and assessment. *The Journal of Technology, Learning and Assessment*, 1(1), 1-24.
- Bennett, R. R. (2011, July). *Innovative assessment systems: The role of new technology*. Presentation at the International Computer Assisted Assessment Conference, Southampton, England.
- Bridgeman, B. (2009). Experiences from large-scale computer-based testing in the USA. In F. Scheuermann, & J. Björnsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. European Commission Joint Research Centre Institute for the Protection and Security of the Citizen. Retrieved from <http://www.jrc.ec.europa.eu>
- Brooks, B. (2009, September). *Harnessing technology to create more holistic assessment and reporting of candidate performance*. Paper presented at the International Association for Educational Assessment Conference: Assessment for a Creative World, Brisbane, Australia.
- Coniam, D. (2009). A comparison of onscreen and paper-based marking in the Hong Kong public examination system. *Educational Research and Evaluation*, 15(3), 243-263.
- Crisp, V. (2008). Exploring the nature of examiner thinking during the process of examination marking. *Cambridge Journal of Education*, 38(2), 247-264.
- Davies, B., & Gralton, T. (2009, September). *Trial of automated essay scoring: New directions for national assessment in Australia*. Paper presented at the International Association for Educational Assessment Conference: Assessment for a Creative World, Brisbane, Australia.
- Dermo, J. (2009). e-assessment and the student learning experience: A survey of student perceptions of e-assessment. *British Journal of Educational Technology*, 40(2), 203-214.
- Dillon, A. (1994). *Designing usable electronic text*. London: Taylor & Francis.
- Fielden, K., & Malcolm, P. (2007). Cell phones in New Zealand schools: Boon, banned or biased. In D. Parsons & H. Ryu (Eds.), *Proceedings, Mobile Learning Technologies and Applications (MoLTA) 2007 Conference*.
- Flexibility, Innovation must guide implementation of new state assessment systems to measure mastery of common state standards*. (2011). Open letter published on the Innosight Institute's website. Retrieved from www.innosightinstitute.org.
- Fluck, A., Pullen, D., & Harper, C. (2009). Case study of a computer based examination system. *Australasian Journal of Educational Technology*, 25(4) 509-523.
- Greatorex, J. (2004). *Moderated e-portfolio evaluation*. Retrieved from http://www.ocr.org.uk/OCR/WebSite/Data/Publication/E-Assessment%20Materials/Moderated_82372.pdf
- He, Q., & Tymms, P. (2005). A computer assisted test design and diagnosis system for use by classroom teachers. *Journal of Computer Assisted Learning*, 21, 419-429.
- Hermans, P. (2009, September). *Developing an e-assessment environment for the 21st century*. Paper presented at the International Association for Educational Assessment Conference: Assessment for a Creative World, Brisbane, Australia.

- Johnson, M., Nadas, R., Bell, J. F., & Green, S. (2009, September). *Marking essays on screen: An investigation into the reliability of marking extended subjective texts*. Paper presented at the International Association for Educational Assessment Conference: Assessment for a Creative World, Brisbane Australia.
- Kelly, P. (2001, April). *Computerized scoring of essays for analytical writing assessments: Evaluating score validity*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, Seattle, Washington.
- Kimber, K., & Wyatt-Smith, C. (2009, September). *Rethinking quality assessment for 21st century learning: How students use and create knowledge online*. Paper presented at the International Association for Educational Assessment Conference: Assessment for a Creative World, Brisbane, Australia.
- Kozma, R. (2009). Transforming education: Assessing and teaching 21st century skills. Assessment call to action. In F. Scheuermann & J. Björnsson, (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. European Commission Joint Research Centre Institute for the Protection and Security of the Citizen. Retrieved from <http://www.jrc.ec.europa.eu>
- Manafy, M., & Gautschi, H. (2011). *Dancing with digital natives: Staying in step with the generation that's transforming the way business is done*. New Jersey, USA: Cyberage Books.
- Moe, E. (2009). Introducing large-scale computerised assessment: Lessons learned and future challenges. In F. Scheuermann & J. Björnsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. European Commission Joint Research Centre Institute for the Protection and Security of the Citizen. Retrieved from <http://www.jrc.ec.europa.eu>
- Partnership of Readiness for College and Careers website. (2011). Retrieved from www.parcconline.org
- Prensky, M. (2001). Digital immigrants, digital natives. *On the Horizon*, 9(5).
- Rallo, R. (2011, July). Is the laptop dead? *New Zealand NetGuide Magazine*, 14-22.
- Scheuermann, F., & Björnsson, J. (Eds.). (2009). *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. European Commission Joint Research Centre Institute for the Protection and Security of the Citizen. Retrieved from <http://www.jrc.ec.europa.eu>
- Smarter Balanced Assessment Consortium website. (2011). Retrieved from www.k12.wa.us/SMARTER
- Stewart, W. (2011, 25 February). Pen and paper must go, says Ofqual head. *Times Education Supplement*.
- The Danish National Assessment System: A case study*. (2011). Cisco Systems 2011. Retrieved from www.cisco.com
- Van Lent, G. (2009). Risks and benefits of CBT versus PBT in high-stakes testing: Introducing key concerns and decision-making aspects for educational authorities. In F. Scheuermann & J. Björnsson (Eds.), *The transition to computer-based assessment: New approaches to skills assessment and implications for large-scale testing*. European Commission Joint Research Centre Institute for the Protection and Security of the Citizen. Retrieved from <http://www.jrc.ec.europa.eu>
- Wesch, M. (2011). *Rethinking education: A new Michael Wesch video*. Retrieved from www.openculture.com/2011/02/michael_wesch_rethinks_education.html

Exploring Four Self-Assessment Strategies to Sustain Learning in Tertiary Education

Roseanna Bourke
Victoria University of Wellington

Email: roseanna.bourke@vuw.ac.nz

Abstract

Self-assessment of student learning at tertiary level has largely concentrated on the notion of self-grading, where students are taught to use a matrix to assign a grade to their work based on their own developed, or teacher-initiated, criteria. Another application of self-assessment is to support students to develop an understanding of their learning, and to encourage them to link such understanding to their work-based or real-world contexts. In this way, students can develop skills to establish their own criteria in a novel context in order to assess their learning outside the confines of course prescriptions and requirements. This paper examines strategies to prepare students to self-assess their learning as a lifelong process, by using four self-assessment strategies (synthesise, evaluate, link, and focus) through online activities that challenge students to examine their learning. Alongside traditional assignments, and block course meetings, four mandatory online self-assessments are introduced during one trimester. A selection of student responses is examined to explore how students interpreted and used these activities to reflect on their learning. The results show that while students use material from the course and reflect on their learning in relation to course content, they can also use self-assessment as a means to inform or reconceptualise their world, to 'look forward' as an active process, and to pose questions as a reflective form of self-assessment. Self-assessment builds on other forms of assessment to support students to assess their learning outside of the course requirements.

Introduction

While the assessment of students in tertiary education fulfils multiple purposes, including certification and credentialing, one critical function is to sustain students' lifelong learning by creating learning opportunities through assessment for them to change their understanding of their own learning (Sadler, 1989). This has increased the use of formative assessment to balance the predominant use of summative assessment approaches in tertiary education contexts. Formative assessment requires students to understand their learning goals, criteria for assessment, and the ability to increase their performance to closely match the desired performance (Sadler, 1989); one strategy to do this is self-assessment (Tan, 2009).

Using Boud's goal for 'sustainable assessment', assessment processes at tertiary level need to enable students to continue their learning (Boud, 2000; Boud & Falchikov, 2005). This research aims to provide a practical application of the use of self-assessment by highlighting four ways that self-assessments can be incorporated into tertiary education courses, and examines how students interpreted and used these activities.

Traditionally in a tertiary environment, summative assessment has been key to the certification and credentialing functions (e.g., through examinations, assignments, essays, and other student-produced outcomes that are 'measured' in various ways), but increasingly the formative nature for supporting learning is being recognised as a legitimate and valid form of assessment (Boud & Falchikov, 2007; Tan, 2007, 2009). Biggs (1998) had earlier argued that if assessment is to *change* students through learning, then formative assessment is necessary alongside summative assessment, although he stated at the time "in most classrooms, even in tertiary classrooms, self-assessment is rare" (p. 104). Since then, formative assessment, and increasingly self-assessment in tertiary education, have been incorporated into the assessment systems with varying

degrees of success for the learners' experience (Boud, 2000; Ibabe & Jauregizar, 2010; Nulty, 2010; Orsmond, Merry, & Callaghan, 2004). The concept of self-assessment acknowledges learning and assessment as complementary processes occurring simultaneously for the student, and studies have shown that self-assessment can raise student achievement (e.g., McDonald, 2007).

For students this means *understanding* what they have learned, what they need to learn, and how to apply this learning in settings outside of their course requirements, and as a process "self-assessment is a sine qua non for effective learning" (Black & William, 1998, p. 26). The research of Boud and Falchikov (2007) and Tan (2007, 2009) argues that self-assessment activities need to be designed to promote the notion of lifelong learning for students. They argue that tertiary educators' role is to support students to successfully problem-solve outside of identified course requirements and to draw on their knowledge, skills, and understanding in order to apply their learning to any given context. In other words, students should be able to continually assess their own ability and performance in relation to the task during, before and after learning takes place. If this is to happen, tertiary educators need to enable these learners to develop the skills to judge themselves, their learning, and their work outside the traditional indicators of marks, grades and externally imposed criteria. Instead, the role of tertiary educators is to support students to function effectively in a constantly evolving technological and social society so that these students have skills and capabilities to tackle future 'unknown' or novel experiences (Bowden & Marton 2003; Nulty, 2010). A key rationale for self-assessment then is that it not only encourages lifelong learning, but also enables students to problem-solve and prioritise their knowledge in novel situations. The following scenario provides an example.

In 2009, a jet airliner with 155 people on board experienced a double bird strike while taking off from La Guardia airport in New York. Less than a minute after take-off, with the plane losing thrust in both engines, the pilot had to make a split-second decision on how to land the crippled plane, and did so through a controlled landing on the Hudson River. The pilot, Chesley (Sully) Sullenberger, was hailed a hero for saving the lives of those on board through his skilful manoeuvring of the plane in what was described as "life-or-death pressure."¹ In an interview with Jon Stewart on the Daily Show in the States, Captain Sullenberger later explained that his decades of flying time "practising to handle whatever might come along" contributed to the problem-solving required at that time. The crew were not specifically trained for such an event, and none had ever needed to undertake such critical and instant decision making in such a situation.

As with any profession, there are times when a decision needs to be made under urgency in situations in which a person has neither had experience nor been 'trained' to do. In such situations the skills of quick reflection, understanding what knowledge to bring to the fore, identifying the priorities for that moment and that situation, and recognising the skills others bring, requires sophisticated levels of self-assessment. This ability to assess beyond course requirements, without reinforcement of grades, and utilising the knowledge of self and others needs practice and support. This paper examines how self-assessment exercises can be introduced at tertiary level to enable students to assess outside course requirements, and to recognise and develop their own ability to self-assess for situations beyond which they are not specifically 'trained'.

Background to Self-assessment

In this section, a discussion on what constitutes formative assessment is necessary to place self-assessment in the context of this research. A distinction is made between self-grading and self-assessment, and links to the learning literature on how self-assessment

¹ Reuters (2009) <http://www.reuters.com/article/idUSTRE50E8AI20090116> [downloaded January 20, 2011]. <http://www.thedailyshow.com/watch/tue-october-13-2009/chesley-sullenberger> (interview with Jon Stewart on the Daily Show).

contributes to learner self-regulation, and understanding of their own learning. The use of summative assessment is recognised as holding a critical and pertinent function in tertiary education but the focus of this paper is to examine how formative methods (through self-assessment) can be used as a complementary process.

Formative Assessment

The distinction between summative and formative assessment lies in the purpose of the assessment. Sadler (1989) argued that three principles underpin formative assessment: (1) students need to know the learning goal; (2) students must have the ability to recognise and make comparisons between their actual and potential performance; and (3) the students must have the ability to act on this comparison. Therefore, the students' understanding of their learning is pivotal to the assessment process. For some, there is the belief that explicit criteria are necessary for students to make their own judgements (Kirby & Downs, 2007), while in contrast, Harlen and James (1997) argue the type of feedback for students will depend on individual students and the context, and therefore no assessment criteria are necessary for the assessment to be formative.

Boud and Falchikov (2005) identify five principles to consider when developing assessment of course content at tertiary level, and each of these relies on a formative process, in addition to any summative function. These principles acknowledge the influence that assessment practices have on the way students learn, and what they choose to focus on for their learning. They include:

1. Assessment "should reflect the learning that is the intended outcome from the program" (p. 3). This recognises the role assessment plays in students' priorities for their learning.
2. Assessment that is designed to be a 'measure of achievement' should come second to the "consequences for student learning" (p. 3). This means that any assessment task that is detrimental to student learning is not educationally valid.
3. "Assessment needs to be demystified if students are to become confident to understand and use it in the world of practice" (p. 4). The authors believe assessment should contribute to the student's ability to self-assess.
4. Students need the opportunity to practise their skills of self-assessment.
5. Assessment and learning tasks should be integrated in order to support long-term learning.

Definition of Self-assessment

Self-assessment is a process that has a range of definitions capturing the use of criteria, grades and student involvement. The most commonly used definition is that of Boud (1991) who defines self-assessment as "the involvement of students in identifying standards and/or criteria to apply to their work, and making judgements about the extent to which they have met these criteria and standards" (p. 5). This definition highlights the importance of criteria for students to 'measure' their work, rather than their learning, and is useful for the notion of self-grading (McDonald, 2007). However, as Eva and Regehr (2008) note, forms of self-grading self-assessments are summative and miss the point of "trying to understand the role of reflection on practice as a pedagogical strategy for better understanding the world around us" (p. 18). Another form of self-assessment at tertiary level is the use of Self-Assessment Modules which are used as a pre-course quiz for students to use online, in their own time, to ascertain their level of knowledge of the proposed course content (Peat, Franklin, & Lewis, 2001). This was an option for students to voluntarily use, and the authors noted not all students chose to access this resource which is of interest given the debate over whether self-assessment in tertiary courses should be mandatory (e.g., Leach, Neutze, & Zepke, 2001; Tan, 2007).

Boud also noted, however, that self-assessment contributes to a student's metacognitive ability, and this is an important move away from increasing concern about the effects of self-assessment that encourage "guess the grade" (Eva & Regehr, 2008, p. 18). Using an alternative view of self-assessment, Mok, Lung, Cheng, Cheung, and Ng (2006) focused on the metacognitive aspects of understanding learning and found that self-assessment reflections prior to, during, and after a learning event contributed to a learner's metacognitive skills. This research uses this alternative approach through four strategies that aim to capture the intentionality and self-regulation of the learner and the social context within which self-assessment occurs. Hence in this view, self-assessment involves the:

deliberate, intentional drive by the learner to calculate the task to be completed, the learning focus needed and the current knowledge that contributes to the goal. It is not possible to self-assess without knowledge of ourselves – that is, of our own identity. Because identity is formed in relationship to others, self-assessment is a social process. (Bourke, 2010, p. 109)

As a means to differentiate self-assessment from self-grading, the focus of this paper is the links to learning: self-regulated and lifelong learning. Recognising that self-grading does have benefits (e.g., Sadler & Good, 2006 identified that self-grading increased student learning, but peer grading did not), this research examines whether students can use self-assessment as a means to reflect on and understand their learning, rather than measure it or become reliant on grades to know they have learnt.

The Role of Self-assessment

One goal for tertiary education is to position students for remaining intentional lifelong learners (Crick, 2007, p. 91). Students need to be equipped to learn beyond university once the network of teaching staff, peers, courses and formal assessment is removed (Boud, 2005, p. 99). The day-to-day dilemmas and activities in the workplace are not posed as examination questions, and necessitate a different skill set to that required of summative assessment in order to excel (McDonald, 2003, p. 3). It has been argued that once learners leave a structured 'learning context,' they may not be particularly effective as lifelong learners when the learning pathway is less directed (Deakin Crick, Broadfoot, & Claxton, 2004). In Deakin et al.'s view, strategic awareness is a necessary and critical component of successful lifelong learning and needs to be encouraged, and this can occur through self-assessment.

Tan (2007) shows how self-assessment is essentially a tool that learners can use to identify and evaluate their own performance outside of "formal study". In a higher education setting, Tan (2007) identified three forms of self-assessment: programme-driven self-assessment, teacher-driven self-assessment, and future-driven self-assessment. From the findings of his study, Tan deemed the third form to be the one best suited to harnessing students' own capacity to make informed decisions and "focus beyond the expectations of the teacher and the programme of study" (p. 120). As one of the students in Tan's study said, "Because it is professionally imperative for nurses, for doctors, for teachers, for everybody, that you're able to realistically self-assess ... *that ability to self-assess is the springboard for your lifelong learning*" (p. 120, emphasis in original). It is the latter form of self-assessment, 'future-driven self-assessment', that this paper aims to explore.

Self-assessment and Learning

The notion that 'sustainable assessment' is an aspect of learning and assessment (Boud, 2000) is consistent with understanding the links between how tertiary learners conceptualise and approach learning. When Marton, Dall'Alba, and Beaty (1993) and Säljö (1979) researched tertiary learners' conceptions of learning, they noted how learners with the least sophisticated and least inclusive conception of learning adopted a surface approach to learning, while those with more inclusive and therefore more sophisticated conceptions adopted a deep approach to learning. These conceptions of learning among

tertiary learners included: (A) Increasing knowledge; (B) Memorising and reproducing; (C) Applying knowledge; (D) Abstracting meaning; (E) Seeing something in a different way; and (F) Changing as a person (Marton et al., 1993; Säljö, 1979). Säljö identified conceptions A, B or C as those most commonly held by students, adopting what he and colleague Marton had earlier described as a *surface* approach to learning (Marton & Säljö, 1976), and conceptions D or E as those held by students adopting a *deep* approach to learning. Other research in this area identified a third, *strategic* approach to learning (Entwistle, 1987). Self-assessment activities can encourage the more sophisticated conceptions of learning such as (E) seeing something in a different way and (F) changing as a person, where learners aim for deep learning and understanding.

Through a study examining learners in a school-based setting (11 and 12-year olds), results showed that students' conceptions of *self-assessment* are also linked to the way they viewed learning, and school and out-of-school contexts generated different conceptions for the same learner. The students adapted and changed for each context and 'learn' accordingly, essentially becoming a 'chameleonic learner' (Bourke, 2010). As shown in Figure 1, the more sophisticated views of self-assessment involved learners seeking the worth of their learning, rather than confirmation that they had learned.

Students' conceptions of self-assessment	
A. Seeking an opinion	Self-assessment is receiving an opinion from an "expert". Students depend on others to confirm that learning has occurred. Teachers and parents are the predominant source of confirmation. This category is characterised by the question: Have I learnt?
B. Getting marks/grades	Self-assessment depends on a symbol (grade, star, stamp, sticker) to confirm learning. For students, these marks and grades identify how well they have learnt in relation to peers. This category is characterised by the question: How much have I learnt?
C. Performing	Self-assessment is viewed as the ability to perform a task. Students use the ability to perform or complete a task as an indicator of learning. They use peers and adults to model the desired performance in order to assess their own performance. This category is characterised by the question: What did I learn?
D. Using criteria	Self-assessment involves the use of pre-established criteria to indicate learning. Students are most likely to view teaching their peers as an indicator of their own learning. Generally, these students believe their learning will improve. This category is characterised by the question: Do I understand what I have learnt?
E. Setting learning goals	Self-assessment is based on setting learning goals. Students are able to set criteria and goals and can evaluate tasks before assessing their own learning in relation to their goals. Students individually self-assess their own learning and use grades as benchmarks for monitoring performance. This category is characterised by the question: What do I want to learn?
F. Evaluating learning content	Self-assessment is part of determining the worth of the learning. Students consider the value of the content and the learning goal before assessing their own learning. They are prepared to persevere with learning if they consider the learning important and valuable. This category is characterised by the question: Is this worth learning?

Figure 1: Students' Conceptions of Self-assessment in a School Setting (Bourke, 2010)

Conceptions of self-assessment that were most sophisticated (e.g., E, setting learning goals, and F evaluating learning content) are more likely to encourage self-regulation in learning which is important for lifelong learning. The links between self-regulated learning and self-assessment have metacognitive and cognitive dimensions: the understanding of assessment standards; the ability to judge the quality of one's work, and to reflect on goal setting and learning progress (Sung, 2010, p. 64).

Self-regulated Learning

Self-regulation involves three areas of psychological functioning: cognition, metacognition, and motivation/affect (Dignath, 2008, p. 80). Combined, these concepts of cognition, metacognition, and motivation enable a learner to self-regulate their learning, and to intentionally draw on what is required to learn, assess, and execute the task required (Boekaerts, 1999; Dignath & Büttner, 2008).

Throughout life, people are faced with new knowledge and skills to learn. Activities such as setting learning goals, planning progress steps, applying pertinent learning strategies, monitoring progress and evaluating learning processes are all part of the self-regulated learning process. In this sense, self-regulation may be seen as “a learner’s competence to autonomously plan, execute and evaluate learning processes, which involves continuous decisions on cognitive, motivational, and behavioural aspects of the cyclic process of learning” (Kistner, 2010, p. 81).

Self-regulation through self-assessment approaches encourages greater self-reliance for the learner and away from the teacher as the sole authority for determining whether they have learned. Through focussing on understanding assessment criteria, learners can develop the ability to critically and dispassionately evaluate their work when making the assessment, which in turn socialises students to assessment processes (Taras, 2008, p. 12). As noted in Figure 1, learning goals can only be achieved if they are understood. Self-assessment tasks are an effective way of systematically developing a learner’s capacity for self-regulation in order to create more structured opportunity for self-monitoring and judging the progression of goals (Nicol, 2006, p. 56).

The aim of this research is to identify the types of practical strategies that can be developed to support students to actively self-assess and increasingly develop skills to self-regulate their learning. By foregrounding four different ways students can self-assess (synthesise, evaluate, link, and focus), the research explores the impact on student learning. The aim is to develop increasingly sophisticated ways to include self-assessment as a legitimate and practical assessment tool for both teaching staff and students in tertiary education.

Method

Students in a postgraduate tertiary education paper, Learning and Motivation, were asked to complete four online self-assessment activities. Each of these self-assessment tasks was presented in a different way and subsequently required the students to engage with their learning in markedly diverse ways to encourage learning beyond the course requirements. Given that Boud and Falchikov (2007) argue for sustainable assessment systems to encourage learners to self-assess beyond course requirements, the four activities in Learning and Motivation used the four strategies of synthesise, evaluation, link, and focus to ask students to think about their learning, their work, and their understanding of the practice-theory nexus.

There were 25 students (15 female, 10 male) enrolled in this postgraduate paper on Learning and Motivation. Many of them were international students (12 countries were represented in this cohort), and while most were from education backgrounds (e.g., working in Ministry or Departments of Education in government, teachers, or specialist teachers) others worked in professions where understanding learning is important (e.g., policing, other government departments, PD consultancy). This research was approved by the Faculty of Education Ethics Committee, and seven students (5 female, 2 male) provided informed consent for their entries to be publicly used for research.

The Self-assessment Exercises

There were four compulsory online self-assessment exercises included in the postgraduate paper in Trimester 2, 2011. Attendance at three two-day block courses, and

the submission of four online self-assessments combined, were worth 10% of the final grade. All students completed the self-assessment exercises fully and at the designated times. The self-assessments were linked to all course objectives. Each self-assessment activity was placed on Blackboard and the students knew the online activity would go 'live' one week prior to the due date. A range of activities was included to provide students with different ways to reflect on their work and to encourage them to demonstrate their learning beyond course requirements. The students had a specified and limited timeframe throughout the trimester to submit their self-assessments online, and once submitted they became 'public' documents within the course. For the purposes of this research, students were asked whether extracts could be taken from their Blackboard entries and used to illustrate how these self-assessment activities were interpreted and used by students. The extracts use no identifying features and students were told that these extracts would not be associated with any identifying features including their gender or grade. The self-assessments will be used to illustrate how students used these activities in their learning.

Self-assessment 1, Synthesis: The first self-assessment exercise was written for students to express their learning through the required readings within the course. Using the text, *Contemporary theories of learning* (Illeris, 2009), students focussed on two chapters for this self-assessment exercise that were designed for them to think about what they had learnt rather than what they recalled.

From the following choices below, what is the most significant thing you have learned from one or both of the readings in your set text. Chapter 8, Biographical learning within the new lifelong learning discourse (Peter Alheit) or Chapter 9, Life cycles and learning cycles (John Heron). This is where you identify (in 250 words) what you learned, how you learned it, how you know, and what aspects you will continue to intentionally aim to learn.

Figure 2: Self-assessment Exercise 1

Self-assessment 2, Evaluate: The second self-assessment activity took place one month later. It provided structure in terms of the reading focus and while it provided choice, all students used both readings.

From the two choices of chapters below (or both) what aspect did you find most relevant to your learning? Chapter 15 A social theory of learning (Etienne Wenger) or Chapter 14 The practice of learning (Jean Lave). This is where you identify (in 250 words) what you learned how you learned it how you know and what aspects you will continue to intentionally aim to learn.

Figure 3: Self-assessment Exercise 2

Self-assessment 3, Focus: The third self-assessment exercise was given to students three weeks after the second and after they had completed a 3,000 word literature review on learning. In the literature review the students had used an extensive literature base to identify key theories (e.g., classical, behavioural, cognitive and sociocultural theories) and conceptual positions (lifelong learning, informal, non-formal and formal learning, communities of practice). They were then asked, through the self-assessment exercise, to define their own theory of learning in 14 words. The 'number' of words was not in itself important; what was critical to this exercise was to support students to capture their notion of learning, in their own words. Given there was a range of students whose English was not their first language, they were told they could express this in their own language and then translate it back to me. However, all students undertook this exercise in English.

Define your own theory of learning to demonstrate your understanding in 14 words exactly.

Figure 4: Self-assessment Exercise 3

Self-assessment 4, Link: The final self-assessment in this course was designed to show students that self-assessment was not always about 'looking back' but that looking forward through questions was an important part in understanding their own learning, and the implications for that learning. Therefore, students were asked to pose a question to the writer of any reading from the course. They were asked to link this question to the writer's intent, and also to their own world. Essentially, the students were to imagine the writer in the room, and identify a question they would ask him/her. The students were asked to provide a rationale for the reading, in part to link the ideas to their own learning and current experience, and in part to demonstrate their understanding of the message.

This self-assessment exercise requires you to ask two questions and give a brief rationale for these. Self-assessment is a process that also involves asking questions in order to learn something for your future context. Using any two of your favourite readings from EPSY 506 (indicate what these are) pose 2 questions directed to the authors of these readings. It is as if you would like them to consider YOUR real world context of learning.

Figure 5: Self-assessment Exercise 4

Results and Discussion

Demonstrated through the online submissions, students began to make links between what they were reading and their own experiences, both as professionals in their fields and learners within the course. This inherently simple shift from seeing themselves merely as a student in this course, towards using the course intentionally as self-regulated practitioners, allowed these students to acknowledge their own learning beyond the course. Their self-assessments show how they qualitatively identify aspects of their learning as a result of the course, rather than, for example, self-grading whether their exercise designated them as a C, B or A grade student. The following two extracts from the first self-assessment exercise show how these students (teachers) intend using their knowledge in their own teaching.

In a knowledge society, the classical teaching-learning model is problematic. Teachers are no longer the 'experts' or 'gurus' of their subject areas. And indeed, they should no longer deem themselves as such. Schools or education institutions should also not see themselves as the only avenue for one to gain learning. I personally feel that sometimes my students seem to learn more during their out-of-school hours than within. The easy access to the Internet has certainly accelerated and enhanced the learning process.

When I am able to see the link between what I am reading to what I have experienced, that's when I learn best. This is especially true when what I read provides some sort of enlightenment to what used to puzzle me or even frustrate me in the classroom. For example, the notion of how learning itself can be learned. This is valuable for me because I can see how I can use this to motivate my students in their learning. If students can be made more aware of what learning actually is about and how learning should not be limited to what they are made to do in schools, and that what they pursue outside school according to their own interests is also valuable learning, then perhaps I can finally help them to develop into lifelong learners who can see that learning is in itself a reward, instead of holding the "what's in it for me" attitude towards learning.

The following examples show how students link their reading to understanding their own experience. In these examples, it is highlighted that whether students agree or disagree with the writer is dependent on how they experienced learning. This provides some insight into how to challenge students to reflect on their own learning. The first example shows a student who expresses disagreement with Alheit (2009) by juxtaposing the concepts against their own experience, and the second example shows another student's views being reinforced through the same reading. Both students read the same text but depending on their position, experienced the reading in their own way.

Alheit reinforced for me my own journey through life and the learning I have encountered on that journey so far. My experiences in a formal school structure were okay but could have been far better if I had been encouraged in a positive and nurturing environment. The notions that learning is life-long and life-wide ring true. I initially learnt far more about real things in life after I had left school, entered the work force and later returned for my structured learning at university. The learning I received outside of school and university was both formal and informal. Informal learning often took place through observing situations or other people and being encouraged to do this by people I knew and worked with.

I disagree with Alheit's view that anyone who never had the chance to learn will not make any effort to acquire new skills late in the life course. As a former Director of Adult and Community Education at [named] College I have worked with older adults whose experiences of learning in life have not been great. However, I believe society is changing and with the advent of new technologies such as cell phones and email I have seen 70 and 80 year olds go to courses to learn about this technology so that they can use it.

The second self-assessment exercise shows how students identify with set readings depending on their own experience and understanding of the topic. To challenge their own thinking they need to grapple with their ideas against those of the writer. Rather than recall what the writer has written, self-assessment requires learners to intentionally create an inner dialogue to self-assess their own understanding. This in turn encourages the learner to reflect more deeply on their learning. As the following extracts show, the second self-assessment exercise challenged the students to think beyond the content. While the first student states initially that the author "seems to say very little", s/he then proceeds to challenge her/his own beliefs about learning and starts to question how learning took place. The second extract shows how a student first situates the reading in a body of literature and then against their own experience, including their future learning goals.

I found chapter 14 [Lave, 2009] "The practice of learning" difficult to get my head around. I feel it is a justifiably "short" chapter because it seems to say very little, other than the obvious; learning involves changes in knowledge, among systems of activity – social, cultural and historical, involving people in society/societies, continually improvising, and often learning by failure. I have "learned" enough about myself over the years to know the problem here lies with me; I am "failing to learn". Maybe this is because of my "socially positioned point of view"; I was raised to see things in "black and white", "right or wrong", "problem – fix it". Maybe it is because I have read the chapter in various ways now, about four times, and I have created my own "non-learning activity" brought about as a "result of anxiety"; and it has made me feel anxious. Maybe I have learnt something?

Wenger's chapter made sense to me from the start, and for my mind, it is taking on a far more complex issue – challenging how we learn, challenging the thinking about learning. Communities of practice are everywhere and more inquiry around a "framework that considers learning in social terms" is needed in NZ. I learnt that this is another example of how learning really is a "life-wide" and life-long process, and a lot of inquiry is yet to be done to improve how we as a society can learn (and teach) the "formal", in less than formal surroundings. How I know I learnt this is because I have started comparing the learning theories in this book, and some overlap, particularly around the social and cultural influences is standing out for me. Because I am quite "challenged" (motivation?) by chapter 14, I am going to intentionally aim to better understand Lave's chapter on the practice of learning.

The responses from the students in the third self-assessment were inclusive of many aspects of learning, and when shared online, and later at a block course face-to-face, students were able to differentiate between different theories and identify the concepts underpinning each idea. Freed from the scholastic form of writing, and finding the '14 word' challenge added interest, the students were keen to synthesise the essence of learning. Examples describing their theory of learning included:

- a synthesis of meaning-making linking previous understandings and new information within different contexts.
- new awareness within the person, anytime, anywhere, intentionally or not, from a resultant motivation.
- change which comes about through a whole person (mind, body, soul) interacting with society.
- living mindfully, unfolding one's potential and that of the systems to which one belongs.
- the disjuncture between the self-interacting with its life world, motivated to effect change.
- a ubiquitous phenomenon that changes what I know and/or how I know something.

The final self-assessment showed students' willingness to imagine what knowledge would contribute to further their own understanding. Using the knowledge they had gained, they were asked to pose questions to the authors with their (the students') current 'real-world' foregrounded. The issues of conflict, context and contradictions were evident, and interestingly all used work-based contexts when they were asked about real-world context. So rather than view themselves as learners and reflect their own personal learning in other ways (e.g., through judo, swimming, sport, hobbies), in general students, as educators, chose to focus on their own students, their learners.

To Petrone 2010. This is a reading that I enjoyed immensely as it put communities of practice into an informal and totally imaginable context (well perhaps for me as a secondary teacher of rampant skateboarding enthusiasts). In his study, Petrone identifies conflict as an important part of learning; in fact learning was contingent upon it. However, in teaching – or especially so in my school – we actively avoid conflict. Practices such as snaking and heckling would be considered bullying, we concentrate on building safe, conflict-free environments and any time we try to explore any ideological conflict it's done so with kid gloves; therefore I would like to ask Petrone: If conflict were an important part of learning, what level of conflict would you imagine seeing students involved in and learning from in a secondary school?

To Nuthall (2007). Nuthall's work emphasises the importance of others in classroom learning, particularly peers of the learner. To him, peer relationships are "semi-private" (i.e., they're not part of the teacher's overall grand plan of managed work and behaviour); however, he identifies them as crucial in the learning process as much knowledge comes from these peers. In secondary teaching, a lot of emphasis has been placed on co-operative learning (peers and groups) in the last 10 years and it has, in my opinion, not always been as powerful as Nuthall suggests it can be which makes me wonder if that is because it is no longer a private, non-threatening interaction between peers but a teacher facilitated interaction that is a part of the public world of the classroom. Therefore, I would like to ask Nuthall: Is the benefit in learning in peer relationships the fact that it is semi-private and not teacher facilitated?

To Engeström: I am [a specialist teacher] working with contradictory activity systems between organisations, schools, parents, teachers, principals and children regularly. Given that the second principle of activity theory is the multi-voicedness of participants, how do you suggest that the power imbalance between parties be addressed so that participants are able to give a frank account of their perceptions of the contradiction?

To Alheit: You speak of the 'dysfunctionality' of established educational institutions, and suggest schools network with their communities to discover new sites of learning and incorporate social and emotional competencies in learning. In NZ the education curriculum identifies key competencies that should be evident within students. Given that the NZ Government and Ministry of Education require evidence-based data to support teachers' professional judgments for academic areas of the curriculum, what 'measures' would you consider appropriate for the key competencies?

Conclusion

Four distinct strategies to self-assessment were trialled in a postgraduate paper. The students willingly embraced the opportunity to explore their learning progress in ways other than traditional forms of assessment. While they completed major written pieces of work (literature review, case study), they also submitted four online self-assessments ranging from 250 words, to exploring their learning theory in 14 words, after having undertaken a 3,000 word literature review in this area. The range of ways to adopt self-assessment practices included posing questions, reflecting on readings in relation to course content, and in relation to real-world contexts, and challenging students to think about why they agreed or disagreed with an author's stance. To do this, they needed both to understand the message portrayed through text, and clarify their own understanding of the concepts. Many expressed some discomfort with this strategy but they were more willing to tackle these tasks without specific criteria than they were with traditional forms of assessment. This took away the view that the lecturer was the 'expert' and seeking a 'known' answer, as self-assessment is highly individual, created in a social context and reliant on the learner's identity to take shape. As a process, it must be shaped according to the student cohort: "Self-assessment should be contextualised – there is no set of self-assessment questions which can fit all cases" (Mok et al., 2006, p. 431).

It was interesting to note that for this first trial, while students did move outside of course requirements to work contexts to self-assess, they did not self-assess beyond this when asked to relate to their 'real-world' context. In future trials, students will be encouraged to think further into their wider learning worlds, and explore the conceptions through their experiences in other places (e.g., sports, hobbies, crafts, community groups, and the arts). For self-assessment to be effective in a tertiary setting, this small scale study has shown that self-assessment activities need to:

1. use the material from the course
2. allow students the opportunity to reflect on their learning in relation to course content
3. provide students the opportunity to explore their learning outside of the course content, but using the material to inform or reconceptualise their world
4. show that self-assessment is a process that looks forward to what students could do, or need to do, rather than 'looking back' and reflecting on what they *had* done or learned
5. support students to think beyond the text and historical context by using their own experiences and posing questions to their authors.

There is a tension between calling for research to establish what forms of self-assessment can enhance student grades, while at the same time exploring ways to enhance students' lifelong learning beyond course requirements. Grades are indicators of how well a student develops skills, and recalls, understands and applies their knowledge, and this aspect remains an important function of summative assessment. The greater challenge remains for students to achieve greater sophistication and skills in self-assessment where they are strongly placed to self-regulate, assess or weigh up decisions outside of course requirements, often in novel situations.

Boud and Falchikov (2005) identify five principles that assessment practices in higher education should consider: reflect the intended learning; foreground consequences for student learning; demystify assessment; practise self-assessment; and support long-term learning. Self-assessment practices alongside other traditional forms of assessment is a promising concept to promote these principles in an integrated way. In our rapidly changing global society the environment for self-regulated learners within and across multiple contexts knows no bounds, and the assessment environment must keep pace; self-assessment is the most authentic and natural response. It is the skill most likely used when problem-solving at speed in a novel and unique situation, which increasingly challenges individuals and society.

References

- Alheit, P. (2009). Biographical learning—within the new lifelong learning discourse. In K. Illeris (Ed.), *Contemporary theories of learning: Learning theorists ...in their own words* (pp. 116–128). London: Routledge.
- Biggs, J. (1998). Assessment and classroom learning: A role for summative assessment? *Assessment in Education: Principles, Policy & Practice*, 5(1), 103–110.
- Black P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7–74.
- Boekaerts, M. (1999). Self-regulated learning. Where we are today. *International Journal of Educational Research*, 31, 445–457.
- Boud, D. (1991). *HERDSA Green Guide No. 5: Implementing student self-assessment* (2nd ed.). Campelltown: The Higher Education Research and Development Society of Australasia.
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151–167.
- Boud, D., & Falchikov, N. (2005). Redesigning assessment for learning beyond higher education. *Research and Development in Higher Education*, 28, 34–41.
- Boud, D., & Falchikov, N. (Eds.). (2007). *Rethinking assessment in higher education: Learning for the long term*. London: Routledge.
- Bourke, R. (2010). *The chameleonic learner: Learning and self-assessment in context*. Wellington: New Zealand Council for Educational Research Press.
- Bowden, J., & Marton, F. (2003). *The university of learning: Beyond quality and competence in higher education* (2nd ed.). London: Routledge.
- Crick, R. D. (2007). Learning how to learn: The dynamic assessment of learning power. *Curriculum Journal*, 18(2), 135–153.
- Deakin-Crick, R., Broadfoot, P., & Claxton, G. (2004). Developing an effective lifelong learning inventory: The ELLI Project'. *Assessment in Education: Principles, Policy & Practice*, 11, 247–272.
- Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students: A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, 3(3), 231–264.
- Engeström, Y. (2009). Expansive learning: Toward an activity-theoretical reconceptualisation. In K. Illeris (Ed.), *Contemporary theories of learning: Learning theorists ...in their own words* (pp. 53–73). London: Routledge.
- Entwistle, N. (1987). A model of the teaching-learning process. In J. T. E. Richardson, M. W. Eysenck, & D. W. Piper (Eds.), *Student learning: Research in education and cognitive psychology* (pp. 13–28). Milton Keynes: Society for Research into Higher Education & Open University Press.
- Eva, K. W., & Regehr, G. (2008). "I'll never play professional football and other fallacies of self-assessment". *Foundations of Continuing Education in the Health Professionals*, 28(1), 14–19.
- Harlen, W., & James, M. J. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education*, 4(3), 365–380.
- Ibabe, I., & Jauregizar, J. (2010). Online self-assessment with feedback and metacognitive knowledge. *Higher Education*, 59, 242–258.
- Illeris, K. (Ed.). (2009). *Contemporary theories of learning: Learning theorists ...in their own words*. London: Routledge.
- Kirby, N. F., & Downs, C. T. (2007). Self-assessment and the disadvantaged student: Potential for encouraging self-regulated learning? *Assessment & Evaluation in Higher Education*, 32(4), 475–494.
- Kistner, S., Rakoczy, K., Otto, B., Dignath-van Ewijk, C., Büttner, G., & Klieme, E. (2010). Promotion of self-regulated learning in classrooms: Investigating frequency, quality, and consequences for student performance. *Metacognition and Learning*, 5(2), 157–171.

- Lave, J. (2009). The practice of learning. In K. Illeris (Ed.), *Contemporary theories of learning: Learning theorists ...in their own words* (pp. 200–208). London: Routledge.
- Leach, L., Neutze, G., & Zepke, N. (2001). Assessment and empowerment: Some critical questions. *Assessment and Evaluation in Higher Education*, 26(4), 293–305.
- Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I. Outcome and process. *British Journal of Educational Psychology*, 46, 4–11.
- Marton, F., Dall’Alba, G., & Beaty, E. (1993). Conceptions of learning. *International Journal of Educational Research*, 19(3), 277–300.
- McDonald, B. (2007). Self Assessment for understanding. *Journal of Education*, 188(1), 25–40.
- McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice*, 10(2), 209–220.
- Mok, M. M. C., Lung, C. L., Cheng, D. P. W., Cheung, R. H. P., & Ng, M. L. (2006). Self-assessment in higher education: Experience in using a metacognitive approach in five case studies. *Assessment & Evaluation in Higher Education*, 31(4), 415–433.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218.
- Nulty, D. D. (2010). Peer and self-assessment in the first year of university. *Assessment & Evaluation in Higher Education*, 36(5), 493–507.
- Nuthall, G. (2007). *The hidden lives of learners*. Wellington: NZCER Press.
- Orsmond, P., Merry, S., & Callaghan, A. (2004). Implementation of a formative assessment model incorporating peer and self-assessment. *Innovations in Education and Teaching International*, 41(3), 272–290.
- Peat, M., Franklin, S., & Lewis, A. (2001). A review of the use of online self-assessment modules to enhance student learning outcomes: Are they worth the effort of production. In K. Kennedy, M. Keppell, C. McNaught & T. Petrovic (Eds.), *Meeting at the crossroads. Proceedings of the 18th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education* (pp. 137–140). Victoria, Australia: ASCILITE.
- Petrone, R. (2010). “You have to get hit a couple of times”: The role of conflict in learning how to “be” a skateboarder. *Teaching and Teacher Education*, 26, 119–127.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144.
- Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11(1), 1–31.
- Säljö, R. (1979). *Learning in the learner’s perspective. I. Some common-sense conceptions*. Göteborg: Göteborg University.
- Sung, Y.-T., Chang, K.-E., Chang, T.-H., & Yu, W.-C. (2010). How many heads are better than one? The reliability and validity of teenagers’ self- and peer assessments. *Journal of Adolescence*, 33(1), 135–145.
- Tan, K. (2007). Conceptions of self assessment: What is needed for long-term learning? In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education: Learning for the long term* (pp. 114–127). London: Routledge.
- Tan, K. (2009). Meanings and practices of power in academics’ conceptions of student self-assessment. *Teaching in Higher Education*, 14(4), 361–373.
- Taras, M. (2008). Issues of power and equity in two models of self-assessment. *Teaching in Higher Education*, 13(1), 81–92.
- Wenger, E. (2009). The social theory of learning. In K. Illeris (Ed.), *Contemporary theories of learning: Learning theorists ...in their own words* (pp. 209–218). London: Routledge.

How do Students use Self-assessment to Learn in a Tertiary Institution?

Roseanna Bourke and Carolyn Tait
Victoria University of Wellington

Email: roseanna.bourke@vuw.ac.nz; carolyn.tait@vuw.ac.nz

Abstract

Self-assessment can be a powerful learning tool for developing the learner's own identity and bridging learning and assessment in tertiary education and real-world contexts. Although self-assessment is not commonly used in higher education, this paper examines its effects on students enrolled in an undergraduate course, *Understanding behaviour: Working with people*. The aim was to understand how compulsory self-assessment can contribute to the learning of tertiary students who are working in professional occupations. Sixty two students chose to participate with each contributing three (n = 3) 250-word self-assessments of learning. These were coded to establish categories and identify themes. Preliminary findings show that students used self-assessment as a tool to analyse self, to understand others, and to understand their learning. Through the use of reflection, self-assessment is a powerful means to support ongoing lifelong learning, and to sustain assessment strategies outside of a specific course content. As an assessment strategy, it has implications for deepening students' learning by allowing them to identify applications of learning outside of a course.

Introduction

Students in tertiary education programmes that are undertaken concurrently while working in their profession often have difficulty linking their *real world* of work to the more theoretical, conceptual studies undertaken through their university programme. Assessment practices within universities, combined with prescribed curricula, encourage certain forms of mechanical responses to essay-type questions, references to a particular literature base and, at times, opportunities to self-reflect on their understanding. University staff can face tensions when, on the one hand, they actively support students through their course of studies which can be prescriptive in content and with pre-determined assessment tasks, while on the other hand encouraging independent thought and action (Boud & Falchikov, 2007).

Assessment purposes are not always aligned and consequently can lead to students choosing to 'play the game' by reporting on, and responding to, assessment requirements that are institutionally valued, graded and that provide their ticket to their qualification (Bourke, 2010; Ramsden, 1988). In other words, students become 'test wise' (Smith, 2011). This paper reports on a distance course that relies on written assessment tasks comprising traditional essay-type questions and case-study responses. In an attempt to support students to negotiate more meaningful engagement with academic text, and to encourage them to connect this learning to their everyday work, a compulsory self-assessment exercise was attached to each assignment to encourage students to reflect on their learning in relation to their work. These were undergraduate students primarily from the New Zealand Police studying towards their Certificate of Contemporary Policing, while the remaining students were from teaching or education backgrounds.

Self-assessment on academic and personal knowledge is premised on 'knowing thyself', and having an awareness of one's own knowledge and the limitations of that knowledge. Undoubtedly, it requires more than this; a sense of vision with an ability to appreciate a wider or more encompassing knowledge-base is critical, and self-assessment is enhanced when others provide feedback on an individual's summary of their perceptions. Reflection in this frame is the ability to use that self-assessment to formulate some action and to act

on those self-assessments. Unless acted upon, self-assessment becomes a summative process, and thus negates the potential of it being a formative process to support learning (Black & Wiliam, 1998; Tan, 2007).

In preparing students for professional practice, the notion of 'self-reflection' or the reflective practitioner, has been promoted as enabling, by giving the student a sense of constant renewal of learning, or learning-in-action, and reflection-on-action (Schön, 1987). Drawing on Schön's earlier work, Mamede, Schmidt, and Penaforte (2008) showed that medical doctors who engaged with self-reflective techniques, increased their diagnostic accuracy of complex conditions.

Although there is no agreed upon definition of reflective thinking (Moon, 2007), in general, it is identified as a process for practitioners to generate informed decisions and to evaluate the outcomes of these (McTaggart & Wilson, 2005). While reflective thinking is central to self-assessment and involves open-mindedness, whole-heartedness and responsibility (Dewey, 1933), the focus tends to centre on the 'problem' or the 'action' rather than the individual. In contrast, self-assessment is designed to focus the learner back onto their own learning in order to understand their role more clearly in that learning, and their subsequent applications of that learning. This article explores how self-assessment was used by students to inform their own learning, and early indicators suggest that the role of reflection contributed to the way in which these students 'accounted' for their learning.

Using Self-Assessment to Inform Learning

Self-assessment can be conceptualised in a range of ways from being a process of 'self-grading' or self-reporting, to one of developing self-knowledge (Peat, Franklin, & Lewis, 2001; Poon, McNaught, Lam, & Kwan, 2009). In this paper, we take a conception of self-assessment that relies on students understanding their learning, and being able to comment on their learning, the aim being to "build the metacognition of learners" (Mok, Lung, Cheng, Cheung, & Ng, 2006, p. 416). In this way, self-assessment as a form of assessment in tertiary education is increasingly being used to encourage student learning and assessment beyond course requirements, particularly those involved in professional programmes of study (Boud & Falchikov, 2007). However, in a tertiary setting, where power and control are systematically and routinely expressed through enrolment procedures, grade averages, assessment requirements and course requirements, and where lecturers experience their own form of assessment expectations within their own institutions, the notion of self-assessment as a sustainable form of assessment (e.g., Boud, 2000; Boud & Falchikov, 2007) becomes a challenging ideal. Although many self-assessment studies have focussed on students' accuracy with their proposed scores and focus on self-grading rather than learning (e.g., Lew, Alwis, & Schmidt, 2010), this paper concentrates on how students use self-assessment to describe and inform their learning, rather than to measure it. When research examines whether student self-assessment is 'accurate' or reliable (e.g., Sitzmann, Ely, Brown, & Bauer, 2010), the focus moves away from how those students use it as a process for their learning.

Even though self-assessment has been linked to developing greater student autonomy and responsibility in learning (e.g., Lew et al., 2010), Tan (2009) outlined the power issues inherent in the introduction of activities such as self-assessment. These power issues arise when lecturers and their students assume that the introduction of self-assessment methods provide more control and power for students in their assessment than might other, more traditional, forms of assessment such as examinations. However, as Tan (2009) argued, "self-assessment may discipline rather than empower students if the ways in which power is exercised over students in self-assessment practices are not understood first" (p. 362). In his work, Tan outlines three forms of power: sovereign, epistemological and disciplinary power bases, depending on how both students and academics perceive and are constrained by power within their institutions or how academics understand and use such

power. These require further consideration because there were specific constraints within the context of this study that self-assessments were introduced. For example, this was a distance paper and submissions could not go online (they were posted or emailed with the assignment); the self-assessment questions had to be pre-determined prior to the start of the semester rather than formulating them during the semester based on student need; and there was minimal variation in that the self-assessments were the same exercise after each assignment. These self-assessments were provided with a mark or grade, and the lecturer had the power over determining that grade. Therefore, students did not have access to establishing their own goals regarding their assignment in order to base their individual self-assessment on those goals. Given that the self-assessment activities were new for the teaching staff, debate among staff focussed on whether or not self-assessments should be mandatory and marked as part of the course requirements.

In his earlier work on differentiating aspects of student self-assessment (teacher-driven self-assessment, programme-driven self-assessment and future-driven self-assessment), Tan (2007) identifies future-driven self-assessment as the ultimate aim for incorporating self-assessment. However, while we agree, for the purposes of this aspect of the course, teacher-driven self-assessment was used and all self-assessments were based on the same question: *What is the most significant thing you have learned from the readings in the module?* The rationale for this question was to ensure students did not move into a self-grading exercise, but rather focused on their self-assessment of their own learning.

Method

The three self-assessment tasks formed the basis of the data analysis. This qualitative study analysed the data without a preconceived framework, and resulted in the development of themes in a manner consistent with a grounded theory approach of allowing themes to emerge from the data (Charmaz, 2006). The data were generated within the parameters of the assessment of a 14-week, 100-level undergraduate course, in trimesters one and two of 2010. Charmaz cautions that this data analysis process requires consideration in terms of its parameters, its purposes and scope, the audience for which it was intended, and what was at stake for the authors (i.e., in this case the students) in producing the text. To counter these concerns, we analysed the data specifically in terms of the course content and for this particular cohort of students in order to learn how self-assessment was used by *these* students.

Students engaged in this distance course individually with their lecturers. There was no provision for online interactive forums or other means of interaction among the students, which arguably impacted on how the students engaged in their learning. For example, it precluded the use of group work and the development of an online community of learners. Therefore, the issue of sovereign power (Tan, 2009) arises because students had minimal input into when and how they submitted their self-assessment, and the pre-determined format provided the authoritative framework. As Tan describes, “in invisible commodity... one is either an agent of power (authority) or a recipient who responds to power (subject)” (p. 362). To counter this, the self-assessment task was employed as a strategy for students to ‘talk’ through their learning in a justifiable and legitimate way through an assessment task.

At the end of the trimester, after the students had received their final grade, they were invited to participate in this study. The students were emailed information about the research and invited to allow the researchers to use their self-assessment tasks in the research. Ethical approval was granted by the Victoria University Faculty of Education Ethics Committee. Students were invited to participate only after the course had finished and grading was completed. This ensured that participation was entirely voluntary and no connection with the grading process could be implied.

Of the 417 potential students, 62 responded and gave permission for the course administrator to access their completed self-assessment, separate them from the assignments, and to assign a number to them to preserve confidentiality. The group participating in the research comprised 48 students enrolled in the Certificate of Contemporary Policing and who held positions in the New Zealand Police, and 14 students enrolled in a Bachelor of Arts in Education.

The majority of the students achieved a B+ grade average for the course. While this study was qualitative and the sampling methods did not aim to be representative, it could be argued that the self-assessment analysis was skewed, given that only those higher achieving students chose to participate. Lew et al. (2010) reported that students' increased ability to self-assess and self-regulate their learning does lead to an improvement in academic achievement, and thus this cohort of students may have particularly valued the self-assessment activity as it enhanced their learning and associated achievements. Therefore, these students may have been more likely to produce richer data by placing value on the self-assessment process. The focus of the analysis became one of identifying *how* these students used the self-assessments to enhance their learning because the researchers did not have ethical approval to compare the self-assessments of lower achieving students.

Each assessment included a self-assessment component of 250 words written in response to the following prompt: *What is the most significant thing you have learnt from the readings in this module?* The assignments were different in nature and content (see Table 1) and the purposes of the self-assessment activities were implicit and not explicitly stated in the course material.

Table 1: Description of Assessments

Assessment	Due date from the start of the course	Tasks	Assessment content focus
1	4 weeks	Short answer response (20%) Self-assessment (5%)	Making sense of theory
2	8 weeks	Essay (30%) Self-assessment (5%)	Working with others
3	14 weeks	Case scenario and analysis (35%) Self-assessment (5%)	Applying knowledge in context

Each student was assigned a number as an identifier. The three self-assessments for each student were integrated into one Word document. The data were then entered into NVivo 8. Each student's data were read by each researcher to generate initial codes from a line-by-line analysis. The researchers compared, discussed and refined the codes using individual examples from the data. Active verbs were used to name the codes. These code names endeavoured to capture the processes that were happening for the students during the self-assessments rather than analysing content themes from the course.

Further coding was then undertaken in an interactive manner and nine of the codes were raised to categories to portray the data. These categories were examined for overlaps and relationships. Where the same data had been entered into two categories the data were reread within the context of the whole of the original self-assessment in order to decide which category appeared to reflect more accurately the meaning of the text extract. These nine categories were then examined for themes.

Results

Nine categories emerged to describe how students used self-assessment in the course and these were sorted into three main themes. Students in this course used the self-assessment activities to analyse self, to understand others, and to understand learning (see Table 2). Each of these will be discussed further in the following sections with indicative quotes.

Table 2: The Use of Self-assessment as a Tool

Analyse self	Understand others	Understand learning
See themselves through the eyes of others.	Understand culture.	Remember course content.
Change as a person.	Identify ideas that challenge thinking.	Link theories to practice.

The results also showed that when students used self-assessment in these ways, they could demonstrate their learning of the course content. When staff on this course met to moderate assignments, and also read the self-assessments of the students, one striking feature emerged. Staff observed how students appeared to articulate their understanding of the course content in a clear way through these self-assessments when freed from the structures of more formal essay-style assignments. Such understanding was not always evident through the formal aspect of the assignment. The students' thoughts and ideas expressed through self-assessment were by nature more personal, and clearly linked to their professional work, but also to the theoretical positions they were introduced to. In contrast, the assignments that were completed in the more traditional style were often stilted, formal, and descriptive rather than demonstrating in-depth analysis or critique of ideas.

Self-Assessment to Analyse Self

One aspect of self-assessment is that, unlike other assessment approaches, it encourages students to learn about themselves as distinct from specific course content. In this research, self-assessment was used by students as a reflective tool to understand themselves as learners, to change themselves, and to see how others might perceive them. In other words, the students used the self-assessment exercises as a means of analysing themselves. Given that the self-assessment exercise was designed to support students to demonstrate their understanding of text, and to increase their engagement with text, this was an unintentional outcome of the self-assessment process. The results showed that students wanted to use the assessment opportunity to learn about themselves.

The first clear example of this category was provided by a range of students who used the self-assessment as a tool to inform their present and previous learning and they invariably situated themselves as learners within these examples. For these students, they looked beyond what they currently 'knew' in order to inform their prior experiences. In this way, a past event (revisited through a self-assessment activity) provided new understandings and became a catalyst for further learning. For example, a student wrote "I personally, have been able to seek out further comprehension of my experiences. I feel like an old wise man, advising a naïve youngster on human behaviour, even though that youngster is me."

In the same way, self-assessment was used to reflect on how individual attributes contribute to learning and working with others. This provided another example of how students used self-assessment to understand their own reaction in a situation. One student examined how the use of non verbal communication was important, and immediately linked this to showing 'emotion'. This was then personally attributed to the way of working:

I have never really paid attention to the way we use non verbal communication. But after reading [the set reading] I realise how effective it can be. I myself often fail to show emotion at times which I realise is like saying nothing. I sometimes think I have developed a sense of emotional hardness overtime during my career in the Police.

Self-assessment spotlighted the effort required to be deeply reflective. The process of reflection was often future focused and impacted on personal goals. For some students, this meant articulating a possible change process. Self-assessment was a means for students to link personal and professional identity, and this provided impetus for changing themselves. In some cases, the self-assessment provided goals for the future:

This means considering different perspectives not only when working with other people but also when viewing myself. By taking a holistic view I will be aware of factors in the inner world and outer world, such as spiritual beliefs and culture. Before approaching and working with people I must acknowledge these factors within myself so I can become aware of any biases I may hold which might affect my relationships with others. I need to recognise all the factors that influence my life and realise that they all work together to form my identity.

The self-assessment provided a means for students to explore the interaction between their learning and their environment as part of analysing themselves. The process of understanding themselves was raised to a theoretical level when students used terms such as 'identity' from the course readings:

Self reflecting on day to day occurrences is actually my identity testing itself within its current environment and that, within itself is a process of learning.

Students analysed themselves using a process of reflection, both on past actions and on future possibilities. This enabled some students to see how other people might view them. Being able to see another point of view in a conflict was a professional skill, especially for the police who could routinely encounter conflict. The self-assessment provided a way of reflecting on how the students' own actions had an effect on others:

I can consciously accept my role in the conflict and can work towards resolving it in a positive way. Before I would have just chalked it up to them being wrong and me being right whereas now it's more of a partnership.

However, this understanding of others was not gained easily as it was founded upon reflection that brought an understanding of self:

It is necessary to understand one's own feelings in order to be able to tune in with and understand other peoples' feelings. This is a thought evoking concept which inspired me to consider how looking inside oneself would not always be an easy undertaking and would often take courage, particularly when facing shortcomings. It made me realise that if we can recognise our own strengths and weaknesses we can better understand, have compassion for and help others who have similar traits.

Self-assessment provided a means of appreciating that there is more than one valid viewpoint. This was both a professional and personal skill. The reflection evident in this theme enabled students to consider not only themselves as individuals, but also how they related to situations in their environment and how others might see them.

Experiences that I once believed to be unproductive or negative are now useful in my ongoing development and help me to connect to others who experience similar problems within their lives. For a long time, I was ashamed of my past and believed that nothing positive would ever come from those experiences. But now I understand that it is through the past that we are able to shape the future.

In this research, we recognised that these self-assessment activities provided a starting point for encouraging students to engage in deeper reflection with the content of the

course. While the readings assigned to the course were connected to the content for each module within the course, the self-assessment activities were used to provide the catalyst for students engaging with the material in ways other than 'remembering' or 'regurgitating' facts. In this way, the self-assessment activities required students to link their knowledge and understanding to their actual work circumstances. The students used the self-assessment activities to analyse themselves within these contexts, which meant that they actively made the course content relevant to their own personal and professional contexts.

Given that the self-assessment exercises reflected the assignment content, and the associated readings, we believe these exercises also provided a means for the students to understand themselves *through* the course content. In this way, the course content is not depicted as knowledge 'out there' to be remembered and recalled; it recognises the interaction between the learner and the course content as being central to the knowledge within each individual.

Self-Assessment to Understand Others

Self-assessment provided a way for the students to explore their understanding of others. This was a complex and multidimensional process. During this process, students compared themselves to others by articulating similarities and differences, which was particularly evident in the dimension of cultural practices. This process of comparison both created stereotypes and challenged them. When referring to Māori, one student observed:

They use different cultural practice using verbal and non verbal language within one another and have very similar values to one another and morals such as having connections with the mountains, tapu, sacredness, spirits, marae welcomes and so on.

Specific aspects of cultural knowledge were of particular interest and were often seen as transferable to professional and personal situations. The course had incorporated readings on Māori culture, history and justice systems, and students frequently used the self-assessments to make connections between their new knowledge from the readings and their workplace to improve their professional or personal skills in interacting with other people. Core cultural concepts from the readings, such as *whānau* and *whanaungatanga*, were linked to a greater understanding of others in both work and recreation. One student saw this understanding of others as being useful "in work or sports, the supervisor or coach." The self-assessment process was seen as a "tool" to make links to workplace skills, such as understanding others, that had the potential to enable the students to do their job more effectively in the future. Students report that the understanding of others' cultural dimensions facilitated rapport, relationships, and negotiation in situations that they might encounter at work. Hence, the self-assessment process opened a space for the students to write about how they might use the knowledge gained from the readings to extend their understanding of others beyond the course.

Another dimension of understanding others related to re-evaluating stereotypes of those people who may be different from the individuals in the course. This was specifically mentioned in relation to values and beliefs. Hence, it was often associated with statements about being less judgemental. In the self-assessment, one student pointed out that "I can appreciate and value that people don't all have the same world view as me and I accept that their world views are just as valid" Expressions such as "[understand] where they are coming from" indicated how students were exploring their understanding of others. Some students used their self-assessments to describe or articulate their thinking about why other people might hold different perspectives to themselves. A student expressed this as "I'm not just dealing with them at this point in time but all their history about how they came to be." This suggested that, while students may have initially been inclined to stereotype, the self-assessment of learning provided a way for students to evaluate their own process of stereotyping through writing.

For some students, this process of understanding others had not only created a new way of appreciating that others viewed the world differently, but also how these views might have developed as different realities came into play for different people.

The realisation that people have such different perspectives on what appears to be the same scenario. However, the perspectives can be so different and people's experiences can be so diverse that it's almost as though the scenario itself changes through different eyes.

The self-assessment data indicated that the process of understanding others was both personal and professional. For some students, it was directly related to knowledge about culture from the readings but other students used their self-assessments to take their learning beyond this and to consider the nature of reality as it might seem for others. The theme of using self-assessment to understand others emerged from the data collected in all three assessment tasks. It was often associated with using self-assessment to understand self and drew on the skill of reflection.

Self-Assessment to Understand Learning

The students used the self-assessment process to articulate an understanding of learning. The data showed evidence of students explicitly recognising their own learning from the course content. Some students made links between theory and their own lives and used the self-assessment process to plan courses of action in the future. When self-assessment was used for reflection on personal and professional learning, it was associated with learning strategies such as endeavouring to apply theory to new situations, connecting with the ideas in the course, and planning for future opportunities to learn.

In many cases, the students used their self-assessment to reflect on particular concepts within the readings that promoted their learning, for example, 'communities of practice'. The imperative to write 250 words in a compulsory self-assessment provided an impetus to move beyond describing the content of the readings. Theory was used as a lens for examining conditions for learning in this example:

Conflicts arise when other communities of practice come into contact with other overlapping communities of practice. In this light, conflicts can be very useful although they are not enjoyable. We need conflict to learn.

The alignment between the goals of the course and the content of the readings was evident when the students used self-assessment to understand their own learning. A common response to the self-assessment prompt was to identify the main ideas from the course readings as the students saw them (rather than as directed by an assignment question) and expressing the learning in their own words. For example, one student wrote, "The readings have provided a nucleus to open our minds to why humans behave as they do." This statement encapsulated the course objectives, allowing the student to articulate his learning and to understand it.

The self-assessment process enabled the students to understand their own learning by providing opportunities for them to evaluate the content of their readings for new learning that could be applied in future in their workplace. This was particularly evident among the Police students who were taking this course for progression in their career. A typical comment in the self-assessment related the course readings to how a student now had a "unique insight" into understanding criminal behaviour.

The process of doing the self-assessment prompted some students to go beyond workplace applications and to write about the interconnected nature of their own learning. They saw their own role in the learning process including their own background and experiences. The process of self-assessment may have assisted students to recognise the complexity of learning.

Realising that, with an understanding of past experiences, we are able to resolve so much when working with others, and finally, of everything that I have studied, seeing the close relationship between all these different aspects come together.

Reflection on the students' own role in learning and the nature of learning in social situations such as in the workplace enabled students to explore the concept of learning.

Students were able to identify ideas that had challenged their own thinking. These challenges were key in stimulating the reflection on the past and action for the future. A common phrase used in these data were "I have come to understand...". One future action reported in self-assessments was a commitment to lifelong learning and an awareness of how that learning might occur. In the case of one student, this came from reading and asking questions to learn from others around him. The purpose of these strategies was to "to better understand my learning."

Within the theme of understanding learning, the data showed students making links to theory and course readings, aligning their learning to course objectives but making it their own by expressing it in their own way, identifying applications within their professional lives, and both reflecting on the past and planning for the future. All of the dimensions of this theme indicated that the students were reflecting on the concept of learning.

Discussion

Each category that arose from the data showed how students foregrounded 'reflection' as their key strategy (as opposed to say, recall) to inform their self-assessment. This was an interesting finding, given the roles that recall, remembering and memory play in other forms of assessment at tertiary level. The process was stimulated by the prompt for the self-assessment which requires the students to look back at their own learning. However, the prompt did not require the students to do anything beyond this. The mandatory self-assessment tasks with the requirement for 250 words created a space which the students could use for their own purposes even though the assessment was teacher-driven. The three themes showing the use of self-assessment identified from the data were influenced by the course content; they also showed how the process of reflecting on course content was used by students for their own professional and personal development.

While Tan (2007) has identified three forms of self-assessment – programme-driven self-assessment, teacher-driven self-assessment, and future-driven self-assessment – this research examined teacher-driven self-assessment with a view to incorporating future-driven self-assessment. Although Tan believed the third form enabled and empowered students to move their focus "beyond the expectations of the teacher and the programme of study" (p. 120), the students in this study were very focused on teacher-driven expectations because the course was pivotal to their gaining their qualifications within the Certificate of Contemporary Policing. However, by the third self-assessment the students realised they were not being 'graded' on the basis of their self-assessment and subsequently became more confident in expressing links between their learning and their everyday work. In this study, perhaps some of their notion of the 'epistemological power' between the students and their lecturers was identified and therefore realised (Tan, 2009). The outcome of this was that some students talked of change and seeing their work in a different way, while others expressed reflective doubt on how they handled a situation and what they would do differently in the future.

Given that the process of reflection builds metacognitive ability (Mok et al., 2006), finding ways to employ assessment practices that encourage reflection at tertiary level is worth exploring. As shown in the students' responses within this research, through providing legitimate opportunities to articulate their understandings of the course without the quality of that process being quantified through a criteria or mark, their increasing ability and preparedness to reflect was evident. If students had been graded based on explicit and

pre-determined criteria of 'worth' for their self-assessment, they may have responded by being more test wise. By allocating a single mark of five points to each completed self-assessment, the students responded by being strategic and doing the self-assessment. Confronted by the need to write an extended paragraph, they turned inwards to reflect so they could justify the significance of what they had learnt for their personal and professional lives. The use of self-assessment in this situation required those teaching on the course to hand over the purposes of the self-assessment to the students. Staff did not have the power to evaluate these purposes and at times, this did create tension.

The three interrelated themes that arose from the data indicated that self-assessment, as used in this study, had the potential to encourage sustainable assessment for these students. Each of the themes – using self-assessment to understand self, to understand others, and to understand learning – show how the learners made the assessment process their own (Boud, 2000). This is an essential aspect of sustainable reflection. These themes suggest that the self-assessment opportunity fostered reflectivity, built learner agency, and constructed active learners, characteristics of assessment tasks which promote lifelong learners (Boud & Falchikov, 2006). This is an area that could be explored in further research, as one of the limitations of the study is that we could not follow-up with these students after the course had been completed to ask them how they used self-assessment as a strategy in their everyday lives.

Conclusion

There are ongoing challenges in a tertiary environment of meeting the requirements for formative and summative assessment, and ultimately providing credentials for students, alongside a rapidly changing work-based environment where influences such as technology and new knowledge create a state of continual evolution. To counter this, in tertiary education we increasingly search to find ways to bridge these two tensions. As shown in this paper, the two are not incompatible as we can extend our capacity to assess student learning by encouraging students to explore their learning beyond course requirements, while drawing on their understanding created through the course content. Early indicators show that self-assessment can contribute to supporting students to engage in their academic work, while developing reflective, social and critical skills for their professional and personal lives.

Through having opportunities to practise self-assessment and self-reflection of learning in a legitimate and acknowledged way in tertiary education, students can be encouraged to think more consciously of the links between course content and their current lives. As one of the students in Tan's (2007) study noted, "it is professionally imperative for nurses, for doctors, for teachers, for everybody, that you're able to realistically self-assess ... *that ability to self-assess is the springboard for your lifelong learning*" (p. 120, emphasis in original). The students began to move towards a reflective notion of learning that expressed future-orientated actions and learning, and has the potential to increase their academic performance. This is consistent with other research in the area, for example, Lew et al. (2010) argued, "as students' metacognitive abilities develop, so do their abilities for self-reflection and self-regulation of learning, which in turn lead to improvements in academic performance" (p. 136). The aim of the self-assessment in this course is to increase students' ability to reflect intentionally on their course content and relate this to their own goals and working experiences. The data in this study indicate that the students actively employed reflection as a strategy to self-assess. Developing self-reflective and self-assessment skills to apply in their professional lives outside of an academic course is exactly what we were aiming for in this course: sustainable assessment in action. Such learning had the potential to bridge the academic, professional and personal worlds of the students.

References

- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7–74.
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151–167.
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 399–413.
- Boud, D., & Falchikov, N. (Eds.). (2007). *Rethinking assessment in higher education: Learning for the long term*. London: Routledge.
- Bourke, R. (2010). *The chameleonic learner: Learning and self-assessment in context*. Wellington: New Zealand Council for Educational Research.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. London: Sage.
- Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. New York: D. C. Heath.
- Lew, M. D. N., Alwis, W. A. M., & Schmidt, H. G. (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education*, 35(2), 135–156.
- Mamede, S., Schmidt, H. G., & Penaforte, J. C. (2008). Effects of reflective practice on the accuracy of medical diagnoses. *Medical Education*, 42(5), 468–475.
- McTaggart, G. L., & Wilson, A. P. (2005). *Promoting reflective thinking in teachers: 50 action strategies*. Thousand Oaks, CA: Corwin Press.
- Mok, M. M. C., Lung, C. L., Cheng, D. P. W., Cheung, R. H. P., & Ng, M. L. (2006). Self-assessment in higher education: Experience in using a metacognitive approach in five case studies. *Assessment & Evaluation in Higher Education*, 31(4), 415–433.
- Moon, J. (2007). *Critical thinking: An exploration of reflective and experiential learning*. London: Routledge Falmer.
- Peat, M., Franklin, S., & Lewis, A. (2001). A review of the use of online self-assessment modules to enhance student learning outcomes: Are they worth the effort of production. In K. Kennedy, M. Keppell, C. McNaught, & T. Petrovic (Eds.), *Meeting at the crossroads: Proceedings of the 18th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education* (pp. 137–140). Victoria, Australia: ASCILITE.
- Poon, W.-Y., McNaught, C., Lam, P., & Kwan, H. S. (2009). Improving assessment methods in university science education with negotiated self- and peer-assessment. *Assessment in Education: Principles, Policy & Practice*, 16(3), 331–346.
- Ramsden, P. (Ed.). (1988). *Improving learning: New perspectives*. London: Kogan Page.
- Schon, D. A. (1987). *Educating the reflective practitioner*. San Francisco: Jossey-Bass.
- Sitzmann, T., Ely, K., Brown, K. G., & Bauer, K. N. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning & Education*, 9(2) 169–191.
- Smith, C. (2011). Examinations and the ESL student: More evidence of particular disadvantages. *Assessment & Evaluation in Higher Education*, 36(1), 13–25.
- Tan, K. (2007). Conceptions of self-assessment: What is needed for long-term learning? In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education: Learning for the long term* (pp. 114–127). London: Routledge.
- Tan, K. (2009). Meanings and practices of power in academics' conceptions of student self-assessment. *Teaching in Higher Education*, 14(4), 361–373.

Two Uses of Value-Added Modelling in New Zealand Secondary Schools

Therese M. Boustead
University of Canterbury

Email: therese.boustead@canterbury.ac.nz

Abstract

Since 1999, many New Zealand secondary and primary schools have used results from value-added modelling to measure and monitor pupil progress through professional self-evaluation rather than public accountability. The Centre for Evaluation and Monitoring (CEM) at the University of Canterbury provides these results using an internationally proven indicator system developed in the UK. Each year more than 100,000 New Zealand pupils, ranging from new entrant 5 year-olds to Year 13 secondary pupils, are registered and measured using a model adjusted to fit the New Zealand school standards-based assessment system. A key component of the model is the ability baseline and its predictive validity. This paper illustrates two ways schools are using the CEM system to identify ways of improving pupil achievement in the early to mid-secondary school.

Introduction

Value-added models in education can range from linear regression analysis (Fitz-Gibbon, 1997) to hierarchical multilevel modelling (Bryk & Raudenbusch, 1992) with the model choice being dependent on precision appropriate to the level of use. The fundamental component of value-added modelling is to determine where a pupil started from at a particular time, called the baseline, and then to longitudinally measure if that pupil kept pace with other pupils who sat that baseline test at the same time under the same conditions. Models such as this may be used to make schools accountable and to compare schools. But, as this paper illustrates, teachers may also use the data in a constructive way to monitor, evaluate and predict the progress of individual pupils, a practice leading to effective pupil evaluation.

The first section of this paper briefly summarises value-added modelling as an indicator and monitoring system. The second section outlines two ways New Zealand schools use value-added modelling for professional self-evaluation, arguing that this approach, as opposed to public accountability, leads to monitoring that focuses on the pupil and not on comparing schools.

Value-Added Modelling

All pupils gain in learning as they progress through a school system, and 'value-added' analysis is defined as a measure of that relative gain. In a nationally-funded British project looking at ways to measure pupil progress, Fitz-Gibbon (1997) concluded that the most valid and reliable approach was through value-added measuring using well-established regression analysis. It was later that the basic regression model expanded into more powerful multilevel models that acknowledged the hierarchical structure of the schooling system and included factors such as pupil background, socio-economic status, school mean and many other factors in addition to prior achievement (Bryk & Raudenbusch, 1992). This paper focuses on the purpose of the value-added measurement and choice of model, the independent variable selected as the baseline and the use of results.

Purpose of Measurement

Countries such as the United Kingdom, United States, Netherlands, Hong Kong, New Zealand and Australia use value-added modelling to measure pupil progress. This modelling can be either government funded or independent. With respect to purpose, there is constant striving for best practice and balance between public accountability and teacher-friendly systems.

Value-added as a Public Accountability System

Value-added models entered the education sector almost 50 years ago from the economics sector as a means of comparing schools and making them accountable for the learning gains (or lack thereof) of their pupils (Gorard, 2006; Schagen & Schagen, 2005). A wide body of research and literature refers to this type of school-related public accountability as 'school effectiveness'. Value-added models were soon extended to include measurement of teachers according to how much progress their pupils made (Kane, Taylor, Tyler, & Wooten, 2010). The outcome of these measures is high stakes for a school and the literature shows that this purpose has more critics of the model than advocates (Amrein-Beardsley, 2008; Gorard, 2006; van de Grift, 2009). In New Zealand, researchers acknowledged that the use of value-added models to define school effectiveness nation-wide had the potential to be costly, elaborate and controversial (Nash & Harker, 1998).

The use of school effectiveness value-added models as a determiner of level of teacher pay and tenure has also led to questioning of the validity of the model and how it is used to assess the extent to which a teacher contributes to pupil learning (Callier, 2010; P. E. Newton, 1997; X. A. Newton, Darling-Hammond, & Haertel, 2010). Teacher contribution is complex and many factors are often not included in these models, such as pupil and teacher choices and teacher decisions (Lockwood, et al., 2007; Misco, 2008; Rivkin, 2007). Likewise, teacher rankings can be affected by pupil characteristics and pupil demographics (X. A. Newton, et al., 2010). High stakes demand a precision that few systems, including sophisticated multilevel value-added models, can deliver. Misinterpretations in multilevel modelling can stem from artefacts such as pupils with initially extreme scores tending to regress towards the mean in later achievement, or initially more able pupils tending to achieve relative higher scores than other abilities, called the Matthew effect (Li, Marsh, Hau, Ho, & Martin, 2005).

However, despite the level of model sophistication, the controversial focus on teachers being accountable for the attainment and progress of their pupils has led the focus of value-added models in a direction that encourages teachers, as part of their professional development, to use value-added models to improve teaching practices and pupil learning (Anderman, Anderman, Yough, & Gimbert, 2010). In addition, simpler models may allow for clearer understanding and interpretation by teachers.

Value-added as a Professional Self-Monitoring System

Using value-added models to improve progress through teacher professional development was an approach proposed about 20 years ago. The approach advocates the use of value-added models to focus on professional self-monitoring by teachers with the aim to improve pupil learning through professional development (Tymms & Fitz-Gibbon, 1995). Professional self-monitoring focuses on performance indicators used by teachers to monitor, target and enhance pupil learning by measuring whether a pupil kept pace with what was expected of his/her. While public accountability is top-down, this approach is bottom-up.

Teachers target for improvement with the aim of improving pupil academic self-concept, pupil performance and pupil achievement. For such professional self-monitoring to be successful, teachers need to have a clear understanding of both the model and the resulting interpretation. This level of clear understanding is not easily attained in the

sophisticated multilevel models. For example, in Hong Kong, value-added monitoring with a focus on non-academic as well as academic self-evaluation was introduced into the secondary system using a multilevel model approach. However, clarity of understanding the model and interpretation of analysis posed a challenge for many secondary teachers, primary teachers, professionals, and administrators (Tse & Albone, 2009).

Although hierarchical multilevel analysis is more powerful, ordinary least square (OLS) regression analysis is easier for teachers to comprehend without losing validity, for there is some evidence that the results from both the multilevel analysis and OLS methods may not differ significantly (Fitz-Gibbon, 1997). Since late 1998, value-added measuring in New Zealand, through the Centre for Evaluation and Monitoring (CEM) at the University of Canterbury (Christchurch), has used the bottom-up approach with the OLS method. Using a value-added indicator system based on work designed in Britain (Fitz-Gibbon, 1997), the CEM value-added focus is on teacher understanding of both the value-added model and interpretation. Causation is not considered or implied in the approach and teachers are encouraged to use the results to diversify their methods as a means of helping their pupils learn.

Baseline as the Independent Variable – Achievement or Ability?

The variety in models used for calculating value-added progress can be compounded by the type of baseline used. For example, progress value-added (PVA) was defined as the gain a pupil makes over and above the gain from pupils with similar prior attainment while concurrent value-added (CVA) was defined as the gain a pupil makes over and above pupils with similar prior ability or position of attainment (Tse & Albone, 2009). PVA is traditionally associated with growth between a prior and current attainment in the same or similar curriculum area while CVA allows for prediction to a wide range of subject areas, is more flexible for comparison both within and between subjects, and is less influenced by prior teaching.

Throughout the world, there appears to be heavy reliance on longitudinal attainment and test scores (PVA) in most of the countries that use value-added models for public accountability such as the United States Tennessee model (Saunders & Horn, 1998). However, countries that emphasise self-evaluation tend to focus on position of attainment or ability (CVA) as the independent variable, such as Hong Kong's Academic Aptitude Test [AAT] (EDB, 2007).

Several evaluation systems are used in New Zealand. Attainment in the curriculum is monitored using national monitoring of a random sample of Year 4 and Year 8 pupils (NEMP) and a curriculum diagnostic system (asTTle) from Year 4 to Year 12. These systems are both government funded. The value-added system conducted by CEM is funded through contracts with individual schools. The baseline is not curriculum dependent, but the outcome measure is. CEM uses a developed ability test as the baseline which is administered under strict examination conditions within a tight time-frame. Consequently, the baseline is administered as pupils enter intermediate or secondary schools from a wide range of schools but pupil curriculum knowledge can be significantly influenced by teaching and learning in the prior school, especially as the level and degree of curriculum taught varies markedly between primary/intermediate schools. CEM found that a curriculum-based mathematics test and an English test did not correlate as well to later curriculum attainment when compared to the CEM developed ability test that combined basic maths skills, numerical thinking skills, vocabulary, visual processing skills, speed and accuracy. Expectation (prediction) is produced by means of a regression line through paired data of the baseline with later curriculum attainment. The advantage of a developed ability test is that it provides a platform for a wide range of later curriculum subjects. While a prior maths attainment score may be a reasonable predictor of current mathematics attainment, it may be a poor predictor of languages or history.

To be included in the feedback from CEM, a subject must meet all criteria which include administration under examination conditions, at least 500 pupils in each subject analysis, correlation coefficients above 0.5 and adequate discrimination with the curriculum outcome with regression slopes greater than 0.2. For a subject to be included in the feedback, pupils who score well on the CEM ability baseline often did well in many curriculum areas. Pupils who do not do so well on the baseline ability test often did not do so well in the curriculum. Analysis is based on real data, and subjects that do not meet all criteria are not included in the feedback to schools.

Pupil Monitoring Using Value-Added Modelling at the Early Secondary School Level

This section shows how value-added monitoring allows schools to effectively target pupil achievement, taking into account where the pupil started from in terms of position of attainment (CVA). The modelling may be used to confirm teacher identification of at-risk and potential underachievers, helping to focus on where improvement may be needed. Value-added modelling can also be used to predict expected achievement in NCEA Level 1, making it possible for every pupil to establish individual goals in specific curriculum areas and allowing schools to enhance the relationship between academic self-concept and academic achievement.

Identifying At-risk and Potential Underachievers

In New Zealand, value-added systems are applied as pupils enter a new phase, whether this is into a new school or into a new assessment phase. Since the largest New Zealand population being measured by the model is at the beginning of the secondary school, the two uses illustrated in this paper will relate to value-added being measured from Year 9 to Year 11, with the baseline administered at the beginning of Year 9 (13 year-olds). In 2011, there were 12,171 pupils throughout the country who sat the 50 minute developed ability test at the beginning of Year 9.

Scenario: A group of boys scored low on a school's entrance test. Under normal circumstances learning for these boys could be slow and teacher expectations low. But in the value-added baseline test of developed abilities the boys scored stronger on their visual processing skills than their other developed ability skills. The school therefore adapted class teaching strategies to capitalise on this strength.

The value-added model used in New Zealand measures whether each pupil kept pace with other pupils of like ability. The key element is to measure the pupil's developed ability rather than curriculum knowledge since the latter can be significantly influenced by curriculum delivery in previous schools or by the introduction of nation-wide curriculum changes. The value-added model in New Zealand uses a series of tests to measure developed ability through a combination of vocabulary, basic numeracy and numerical thinking skills, visual processing, speed and accuracy. Value-added modelling of data collected by the CEM over the past 13 years found that the weighted combination of these skills was a consistently good predictor of later curriculum achievement in a wide range of traditional curriculum areas.

The vocabulary and numerical skills are timed sections of the test, but the visual skills, speed and accuracy are a combination of timed sections. Research has shown that each developed ability skill is linked to success in the classroom. For example, research in the south-eastern United States with second and fourth graders reinforced the link between vocabulary and the quality of narrative writing (Olinghouse & Leaird, 2009).

The results of the value-added ability baseline test are divided into four sections, with an overall score (called the MidYIS score based on work by Fitz-Gibbon (1997)) as the weighted combination of all sections. Each score is standardised to a mean of 100 and standard deviation of 15. This means that scores above 130 are significantly high and scores below 70 are significantly low. The test is not an IQ, but a measure of skills that correlate well with curriculum attainment.

Schools, as in this scenario, use the baseline developed ability data of each pupil to compare with other curriculum-based assessment results being administered at the same time. Schools are able to focus on identifying potential underachievers through inconsistencies between developed ability and curriculum skills, resulting in a focus for further investigation and future close monitoring.

A potential at-risk pupil can be identified by demonstrating significantly low scores on both curriculum content and developed ability. As illustrated in the scenario, often these pupils have already been identified by the teachers and the developed ability assessment can both confirm the at-risk status as well as point to strengths and weaknesses in different developed ability skills (Figure 1).

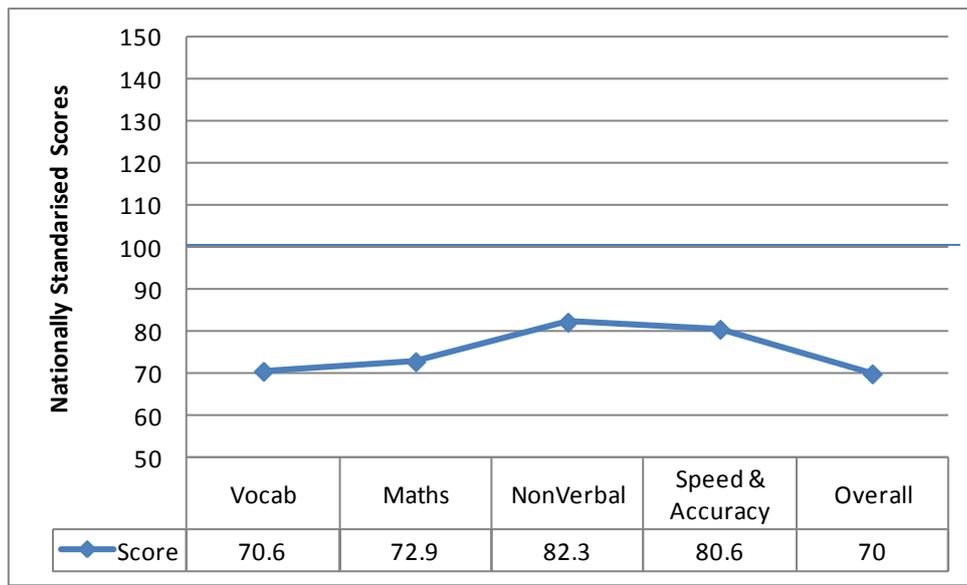


Figure 1: Potential 'At-risk' Pupil with Low Developed Ability Score. Standardised Results Shown for Vocabulary, Mathematics, Non-verbal and Speed/Accuracy

Scores 70 or below are significantly below the group mean. Group average =100, standard deviation =15

In Figure 1, this hypothetical pupil scored significantly low in the developed ability test, but showed slightly higher strength in non-verbal skills, speed and accuracy. Low attainment in curriculum mathematics and literacy reinforces the level, but Figure 1 shows that strategies to help the pupil could include a visual approach.

The focus on at-risk pupils is not limited to low-scoring pupils. An above average pupil as defined by the developed ability assessment (Figure 2) is expected to score well in their curriculum tests.

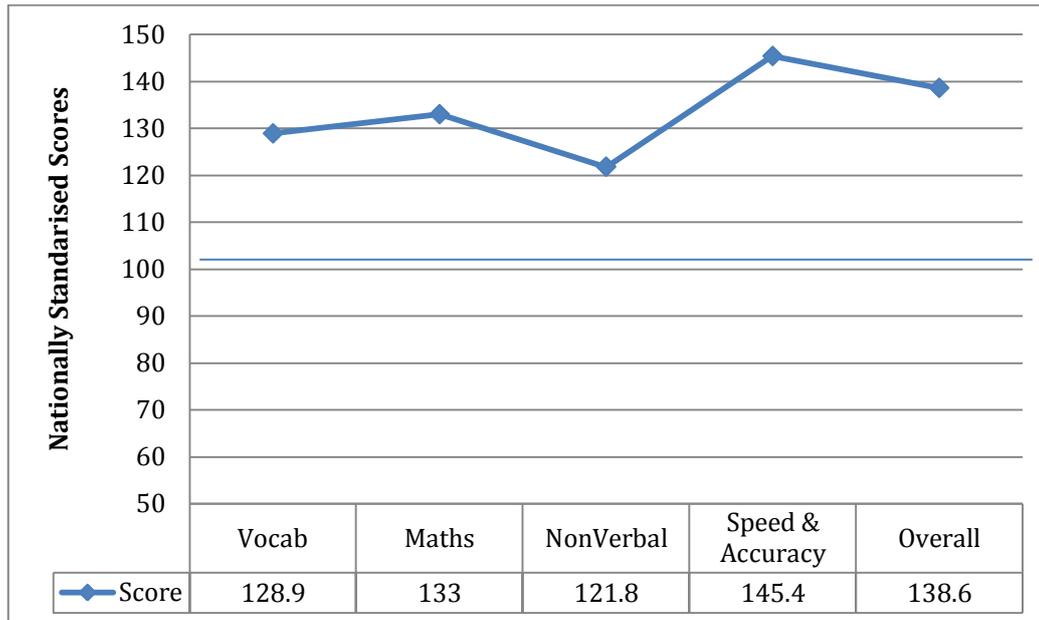


Figure 2: Pupil Baseline Result Showing Standardised Results for Vocabulary, Mathematics, Non-Verbal And Speed/Accuracy for Above Average Developed Ability

Group average =100, standard deviation =15

According to Ogbonni (2009), above average pupils whose abilities do not match attainment include those:

- who do not perform according to expectations in a particular subject area
- who do not show interest / do well in their studies because of behaviour
- who do not perform well in a specific subject area
- who underachieve despite having the necessary ability to do better
- who are limited by cultural, language and gender factors from doing well academically at school.

This definition is broad enough to include most categories of underachieving students, including gifted underachievers whereby highly able pupils perform poorly academically having high potential. An increasing number of New Zealand schools are beginning to use the value-added modelling data to raise the level of pupil achievement of low and high-scoring at-risk pupils.

Targets to NCEA Level 1

Scenario: Teachers were concerned for an able pupil who openly stated that she would do the absolute minimum to pass NCEA Level 1. Consequently her teachers needed to encourage this pupil to achieve to her ability level and, with both the pupil and teachers knowing what her likely scores should be, to focus on improving self-concept and obtaining expected (or better) achievement in each of her subject areas.

Prediction of an expected achievement or outcome is known as a target. Educational target setting can range from individual subject targets to large-scale government directed target setting associated with government-led policy. In England, government-directed target setting has led to controversy over control, pupil and teacher ownership, and the degree of stakeholder involvement since its introduction in 1999 (Flecknoe, 2001; Waite, Lawson, & Bromfield, 2009). The thrust of target setting in this situation is about school effectiveness, that is, value-added modelling as an accountability tool, with the pupil excluded from input (Fielding, 1999; Waite, et al., 2009).

Some schools in England have experienced positive results using targets but this may depend on school success. For example, a sample of 12 United Kingdom schools showed that teachers were more likely to express ownership of the targets, more likely to use a wide range of different targets, believed that their school's information system supported the target process and were more likely to focus on individual pupil attainment if more than 50% of their pupils passed the GCSE examination with C grade or higher (Davies, Coates, Hammersley-Fletcher, & Mangan, 2005). Research has also shown that there is more effective dialogue between pupil and teacher if targets are used alongside academic tutoring (Younger & Warrington, 2009).

Pupil input is a critical factor in the successful use of targets. A frame of reference that underpins the use of target graphs when used to help pupils progress and improve attainment is the Big-Fish-Little-Pond-Effect (BFLPE) (Marsh, 1987; Marsh & Craven, 2006; Marsh & Parker, 1984; Marsh, Trautwein, Ludtke, Baumert, & Koller, 2007). Using a pathway analysis approach, Marsh showed that academic self-concept was different to general self-concept and self-esteem, with academic self-concept being a specific facet of general self-esteem. Marsh (1987) also confirmed the apparent paradox that school ability had an inverse effect on academic self-concept. Pupils attending a high-ability school had lower academic self-concept and were likely to obtain lower scores than an equally able pupil in a low-ability school. This is because the pupil in the higher-ability school compares him/herself with the more able pupils, resulting in a lower academic self-concept. The researchers also found a mutual relationship that "improved self-concepts will lead to better performance and improved performance will lead to better self-concepts" (Marsh & Craven, 2006, p. 159). Not only could BFLPE exist at the late secondary school level, but could be long lasting four years later (Marsh et al., 2007).

In New Zealand, there is increasing interest in individualised target predictions to NCEA Level 1 with pupil input. In the absence of high stakes, this interest comes from individual school initiative rather than direction by government. Consequently, target setting from a self-evaluation perspective is bringing out some of the positive aspects of the model's use.

Targets produced from CEM's value-added model in New Zealand are calculated by importing a Year 9 ability baseline score into the latest subject regression lines formed from the previous year's data. An example of a typical target graph for a hypothetical high-achieving pupil is shown in Figure 3. Each bar shows a separate curriculum area: accounting, biology, history, French, English and mathematics. The height of the shaded area in each bar indicates the proportion of Achievement, Merit or Excellence grades expected to be reached for 20 credits sat in each of the curriculum areas. The three horizontal lines are the thresholds for all credits with an Achieved grade (bottom bar), all credits with a Merit grade (middle bar) and all credits with an Excellence grade (top bar).

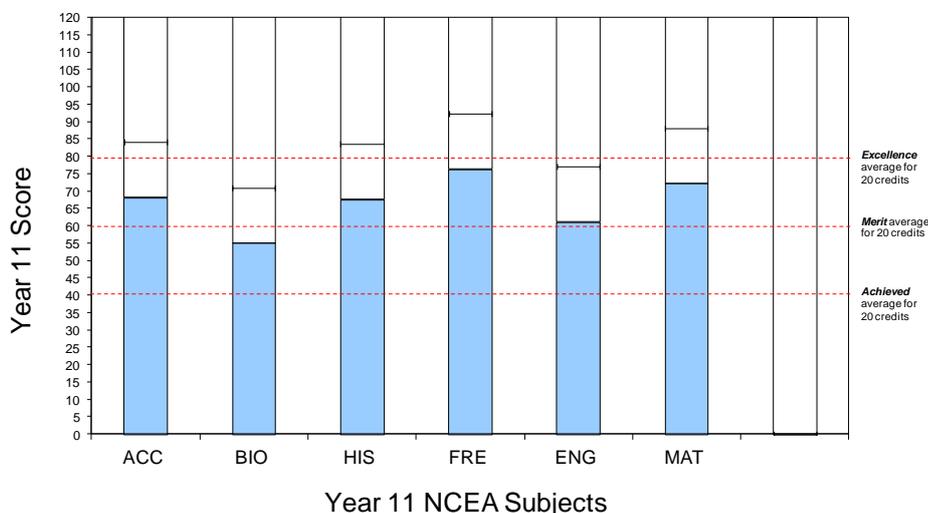


Figure 3: Minimum Targets Calculated for Pupil A to NCEA in Year 11.

At the beginning of Year 11, Pupil A illustrated in Figure 3 is expected to achieve all 20 credits of English with a Merit grade, most of the 20 biology credits with a Merit grade, just over half the 20 mathematics credits with an Excellent grade and the rest with a Merit grade. Most of the French credits are expected to be passed with an Excellence grade.

New Zealand schools are using such targets in a variety of ways to improve academic self-concept and encourage pupils to achieve obtainable goals by adapting teaching strategies to meet the needs of their pupils. Target-setting can be positively embraced when there is flexibility for the schools and staff to adapt targets to their own system (Higham, Sharp, Machin, & Wilson, 2001). A caution is that the targets can depict the minimal level for keeping pace with others of like ability. Highly motivated pupils whose progress is well beyond 'keeping pace' could have targets well below achievable outcomes. Some of this limitation can be overcome by personalising the targets beyond the level expected of the pupil.

However, a better use of target graphs is to focus on pupils who require help to improve achievement. In cases that involve minimum credit counting, as seen in the scenario, a target graph empowers the pupil to self-monitor his/her progress throughout the year by recording step-wise successes in internal achievement on the graph. This use of the predictive nature of value-added modelling focuses on the individual pupil and involves them in setting and monitoring their own progress with teacher support.

Discussion

Value-added modelling measures relative pupil progress and has the potential to be both an accountability tool and a professional self-evaluation tool. Many value-added models are used throughout the world, ranging from simple regression analysis to sophisticated multilevel modelling, the latter dominated by government-driven accountability with a high-stakes competitive school comparison and teacher effectiveness. High-stakes accountability and professional self-evaluation may use the same or similar value-added model, but they have contrary aims and effects. For example, seeking to improve in teaching and learning becomes the focus of professional self-evaluation, in contrast to often unintended outcomes of controversy, mistrust and a desire to cover or manipulate less desirable outcomes engendered in high stakes public accountability.

By introducing a simple regression value-added model many years ago, New Zealand schools have embraced value-added modelling without controversy or mistrust. The New Zealand system is based on research by Fitz-Gibbon (1997) with a focus on professional self-evaluation as opposed to public accountability. The simple regression analysis sacrifices high-stakes detail for indicator systems that are clearer and more easily comprehensible for teachers.

Just as high-stakes multilevel modelling accountability systems are heavily criticised for being costly, controversial and difficult to understand, the use of the simpler regression analysis approach has its limitations. The simpler regression analysis used in New Zealand can be understood by teachers but interpretation needs to take into account that the system is a rough indicator, not a precise measurement tool. The simpler regression analysis model does not take into account the hierarchical structure of the school system, an essential element if comparing schools. The model assumes all schools are equal, irrespective of average ability intake, decile ranking, teaching programmes, and learning outcomes. This may not be the case within large cities. Another limitation is heavy reliance on the baseline as a measure of initial intake, when prior intake may account for only a third of the variance for progress. The baseline is set once, which is a further limitation. A pupil who is unwell or who underachieves on the baseline measure is measured with lower intake ability pupils and, by keeping pace with their true ability counterparts, that pupil will obtain a higher progress score.

To counterbalance some of the limitations, CEM's model has baseline testing conducted under examination conditions in each school within a limited timeframe and most of the baseline tests take place within two weeks of schools commencing the academic year. All pupils sit the same developed ability baseline, an audio controls the timing of each section of the test and correlation coefficients between the baseline and curriculum outcome scores must be good enough to develop a stable regression line for each subject. Any pupil who does not sit the baseline test within the timeframe is excluded from the analysis and any subsequent use of value-added data. Those subjects that do not produce a stable regression line are not provided in CEM's feedback to schools. The regression lines formed from the New Zealand data are positive, do not vary significantly from year to year and only subjects with acceptable correlation coefficients are provided in the feedback to schools. Although comparative analysis has been periodically conducted on the data with multilevel models for research purposes, availability of explanatory variables is limited. With the increasing use of the value-added data for targeting, further research needs to extend into growth mixture modelling to determine paths and trajectories from sub-populations of the data.

The simple regression analysis value-added approach empowers the teacher to analyse their own class pupils and groups for improvement. Two uses that the New Zealand schools make of the value-added model are outlined in this paper. Both these uses arose from schools where teachers not only comprehended how the value-added data were calculated, but could see ways in which the data could meet other needs. There are many other uses, but this paper outlines two uses which came from empowerment of model understanding not only how the value-added model is calculated, but also the limitations of the model.

The first use is comparing the baseline developed ability information with curriculum assessments to identify potential underachievers. Most of the curriculum areas in the analysis have three consecutive years of regression lines that do not significantly differ from year to year. Each regression line is specific to a curriculum area. The developed ability baseline correlates well with at least 18 different curriculum subject areas and this gives support to the assertion that the baseline is a reasonable measure of underlying curriculum skills. The link between developed ability correlating to later curriculum achievement and current curriculum assessment is a way for identifying pupils who do not have the expected curriculum knowledge that requires further investigation.

A second use that came from innovative teachers understanding how value-added is calculated is the increasing use of prior value-added regression equations to develop targets for their pupils. As the pupils reach Year 11, their Year 9 developed ability score is used with the previous Year 11's data to give a rough indication of the minimum expected target for the pupils. More research needs to be done in this area. What is the best way for targets to be used to improve performance? If targets can be used at an earlier age to improve academic self-concept, how can they be used to improve academic self-confidence and attainment? What effect does BFLPE have on targets being given to pupils in a high ability school? Is it wise to give targets to pupils who are at the extreme ends of academic attainment or who are already working well above expectation?

The two uses of the value-added model outlined in this paper are not the only ways schools can benefit through using a good predictive model via self-evaluation. In addition to a focus on the pupil, teachers can monitor their own curriculum area and target groups of pupils who have not performed as expected, or have exceeded expectation. Schools are already using the CEM value-added data in this way.

CEM's value-added system in New Zealand is voluntary. Output data are owned by the school and it is initiatives by the schools that have led to the two uses mentioned in this paper. Teachers are encouraged to use the data to identify underachievers early, adapt their teaching methods and help pupils improve their academic self-concept through individual targeting. These strategies aim to improve achievement outcomes.

Current research points to affective measurements in terms of pupils' attitudes, interests and values as being missing in assessments (Hall, 2011), but a focus on the pupil, early intervention, monitoring using a good predictive model and setting targets or goals goes a long way to embracing affective measures as well as maximising use of what has proven to be a very valuable and useful system for schools.

This paper illustrates how value-added modelling may be used to evaluate pupil achievement so as to maximise class and individual success in secondary schools. The simplicity of the CEM value-added model means that teachers can understand what is being measured, how the measurements are done and what limitations come with the data so that they can discover new ways to use the measurement where it is needed most for their pupils. There is a potential for similar use to be made in primary schools.

References

- Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, 37(2), 65-75.
- Anderman, E. M., Anderman, L. H., Yough, M. S., & Gimberty, B. G. (2010). Value-added models of assessment: Implications for motivation and accountability. *Educational Psychologist*, 45(2), 123-137.
- Bryk, A. S., & Raudenbusch, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.
- Callier, J. (2010). Paying teachers according to student achievement: Questions regarding pay-for-performance models in public education. *The Clearing House*, 83, 58-61.
- Davies, P., Coates, G., Hammersley-Fletcher, L., & Mangan, J. (2005). When 'becoming a 50 school' is success enough: A principal-agent analysis of subject leaders' target setting. *School Leadership and Management*, 25(5), 493-511.
- EDB. (2007). Aims of education and quality indicators (Ch 2). In *Hong Kong Education Commission Report No. 7: Quality school education*. Retrieved from <http://www/edb/govt.hk>
- Fielding, M. (1999). Target setting, policy pathology and student perspectives: Learning to labour in new times. *Cambridge Journal of Education*, 29, 277-287.
- Fitz-Gibbon, C. (1997). *The Value Added National Project final report: Feasibility studies for a national system of value-added indicators*. London: The School Curriculum Assessment Authority.
- Flecknoe, M. (2001). Target setting: Will it help raise achievement? *Educational Management and Administration*, 29(2), 217-218.
- Gorard, S. (2006). Value-added is of little value. *Journal of Educational Policy*, 21(2), 235-243.
- Hall, R. (2011). Affective assessment: The missing piece of the educational reform puzzle. *The Delta Kappa Gamma Bulletin*, Winter, 7-10.
- Higham, J., Sharp, P., Machin, D., & Wilson, M. (2001). Academic progress in the 16-19 curriculum: Some senior staff perceptions of current monitoring systems and operational procedures. *The Curriculum Journal*, 12(1), 59-80.
- Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2010). Identifying effective classroom practices using student achievement data. *The Journal of Human Resources*, 46(3), 587-613.
- Li, X., Marsh, H. W., Hau, K.-T., Ho, I. T., & Martin, A. J. (2005). *Regression toward the mean artifacts and Matthew effects in multilevel value-added analysis of individual schools*. Paper presented at the Australian Association for Research in Education Conference, Parramatta, Australia.
- Lockwood, J. R., McCaffrey, D. F., Hamilton, L. S., Stecher, B., Le, V., & Martinez, J. F. (2007). The sensitivity of value-added teacher effect estimates to different mathematics achievement measures. *Journal of Educational Measurement*, 44(1), 47-47.

- Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology, 79*, 280-295.
- Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects model of self-concept and performance from a multidimensional perspective. *Perspectives on Psychological Science, 1*(2), 133-163.
- Marsh, H. W., & Parker, J. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology, 47*, 213-231.
- Marsh, H. W., Trautwein, U., Ludtke, O., Baumert, J., & Koller, O. (2007). The big-fish-little-pond effect: Persistent negative effects on selective high-schools on self-concept after graduation. *American Educational Research Journal, 44*(3), 631-669.
- Misco, T. (2008). Was that a result of my teaching? A brief exploration of value-added assessment. *The Clearing House, 8*(1), 11-13.
- Nash, R., & Harker, R. (1998). *Making progress: Adding value in secondary education*. Palmerston North: ERDC Press, Massey University.
- Newton, P. E. (1997). Measuring comparability of standards between subjects: Why our statistical techniques do not make the grade. *British Educational Research Journal, 23*(4), 433-450.
- Newton, X. A., Darling-Hammond, L., & Haertel, E. (2010). Value-added modelling of teacher effectiveness: An exploration of stability across models and contexts. *Education Policy Analysis Archives, 18*(23), 1-23.
- Ogbonni, C.-E. (2009). Underachieving learners: Can they learn at all. *Annual Review of Education, Communication and Language Sciences, 6*, 84-102.
- Olinghouse, N. G., & Leaird, J. T. (2009). The relationship between measures of vocabulary and narrative writing quality in second- and fourth-grade students. *Read and Writing. Published online by Springer Science + Business Media B.V. 2008, 22*, 545-565.
- Rivkin, S. G. (2007). *Value-added analysis and education policy*. Washington, DC: National Centre for Analysis of Longitudinal Data in Educational Research.
- Saunders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee value-added assessment system (TVAAS) database: Implications for educational evaluation and research. *Journal of Personal Evaluation in Education, 12*(3), 247-256.
- Schagen, I., & Schagen, S. (2005). Combining multilevel analysis with national value-added data sets: A case study to explore the effect of school diversity. *British Educational Research Journal, 31*(3), 309-328.
- Tse, H., & Albone, S. (2009). A school-based evaluation of reading gains and the use of a value-added grid for marking interpretations of assessment results. *New Horizons in Education, 57*(2), 59-73.
- Tymms, P. B., & Fitz-Gibbon, C. T. (1995). Students at the front: Using performance indicators for professional development. *Educational Research, 37*(2), 107-122.
- van de Grift, W. (2009). Reliability and validity in measuring the value added of schools. *School Effectiveness and School Improvement, 20*(2), 269-285.
- Waite, S., Lawson, H., & Bromfield, C. (2009). Individual target setting in a mainstream and special school: Tensions in understanding and ownership. *Educational Studies, 35*(2), 107-121.
- Younger, M., & Warrington, M. (2009). Mentoring and target-setting in a secondary school in England: An evaluation of aims and benefits. *Oxford Review of Education, 35*(2), 169-185.

Assessing Group Work in Student Industry Projects: Is Fairness Achievable?

Jill Clark, *Whitireia New Zealand*
Trish Baker, *Wellington Institute of Technology*

Email: jill.clark@whitireia.ac.nz

Abstract

International researchers have produced extensive evidence of the benefits of cooperative learning; as a result, New Zealand tertiary educators have been encouraged to incorporate cooperative learning techniques into their programmes. Many undergraduate and postgraduate business degrees now include substantial industry projects where students work in groups to address real life industry issues. Anecdotal and empirical evidence in New Zealand suggests that these projects, with their focus on cooperative learning, can present significant assessment challenges for both lecturers and students. A major concern for both students and lecturers is the perceived unfairness of assessment.

This paper examines assessment methods that are appropriate for long-term group projects in New Zealand tertiary institutions, discusses the implications of assessing both product and process, and recommends assessment strategies to address perceptions of unfairness and social loafing.

Introduction

A search of Google produces 7,540 hits for “I love group projects” and 9,230,000 hits for “I hate group projects.” This result will not come as a surprise to many tertiary lecturers who are required to incorporate group projects into their programmes. The benefits to students of working in groups are well documented. International researchers have concluded that students who learn in groups develop increased intercultural understanding (Aronson, 2001; Slavin, 2001), enhanced interpersonal skills (Johnson & Johnson, 1999), better mastery of complex material and improved long-term retention (Johnson, Johnson, & Smith, 2007). Chen, Donahue and Klimoski (2004) believe that involvement in cooperative learning provides more effective preparation for the modern participative workplace. In their seminal work on cooperative learning Johnson and Johnson (1989) pointed to more frequent use of higher-level reasoning and meta-cognitive thought, more accurate and creative problem solving, more willingness to take on and persist with difficult tasks, more intrinsic motivation, and transfer of learning from one situation to another as beneficial outcomes of effective group work. Davis (2009) summarised the overall benefits of cooperative learning by stating that learning in groups promotes “deep” as opposed to “surface” learning, “active” as opposed to “passive” learning, and encourages learning through experience and problem solving.

A major concern for both lecturers and students, however, is the perceived unfair assessment of work carried out in groups. It can be difficult for lecturers to devise assessment strategies that will address issues of individual and group accountability, will incorporate measurement of process and product, and will address perceptions of unfairness and social loafing (the tendency for individuals to reduce their efforts when working in groups).

Assessment issues are particularly relevant where students are required to work in groups to complete a ‘real life’ project for an industry client. An industry project is a significant piece of supervised study completed over a period of several months, normally in the final year of an undergraduate degree. Students taking such a course are expected to demonstrate both a professional attitude and the ability to integrate the various disciplines acquired during their study. These disciplines include academic knowledge and project

management skills as well as the ability to work effectively in a group to produce outputs that meet industry standards.

This discussion paper aims to assist New Zealand tertiary business lecturers to design assessment methods for group industry projects that will support, not undermine, the benefits of the group experience. It will examine important assessment issues such as freeloading, individual and group accountability, measurement of process and product, and equity, offer a range of assessment methods, and present an example of an integrated assessment design for student industry projects.

Discussion

Industry Projects

Many undergraduate business programmes include an industry project where students work in groups to complete a project for an industry client in the final year of their degree. An industry project is an authentic, project-based activity that closely relates to professional work in the field and is intended to be a culminating educational experience for the students, exposing them to industry standards and expectations. The project provides the opportunity for students to integrate their learning from previous courses, and to demonstrate their development of both discipline knowledge and generic skills. For the institution, it enables assessment of how well the students meet the graduate profile for the programme.

It is expected that students undertaking an industry project will meet a wide range of learning outcomes including the ability to:

- apply both academic and generic knowledge and skills learnt in prior papers on their degree and develop new skills
- plan and manage all aspects of the project
- manage the group
- maintain appropriate and professional communication with stakeholders in the project
- produce professional outputs that meet academic and industry standards
- document and reflect on all aspects of the process and individual learning.

Figure 1 shows the complex mix of product and process elements that make up a student industry project. The product includes the tangible outputs that the group produces for the client and for academic purposes. This may include a project proposal, reports, presentations and data analysis. The process includes management of the project and the group, and individual reflective work. Appropriate assessment must not only address the product and process elements of the project but also must be linked to all the stated learning outcomes.

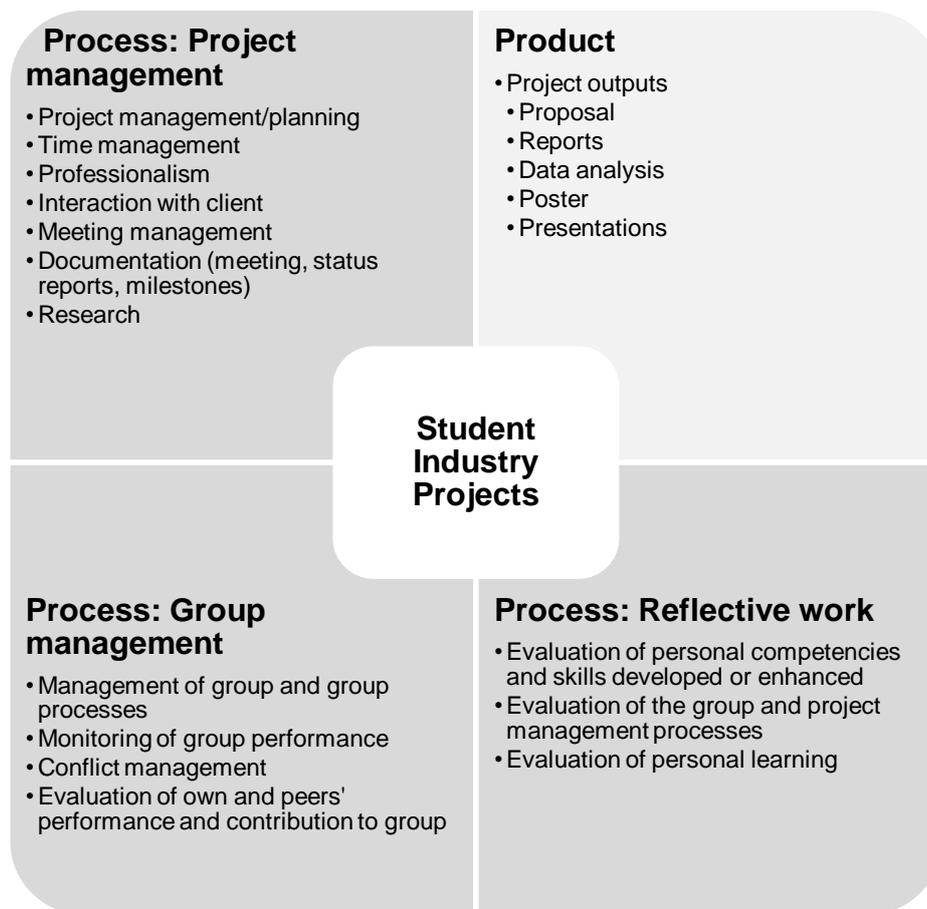


Figure 1: Elements of Student Industry Projects

Assessment Issues

The first step in determining assessment methods for industry projects is to decide what must be assessed as decisions around assessment send clear messages to students about the valued outcomes of the task (Millis & Cottell, 1998). “The use of explicit assessment criteria has the benefit of making all students aware of what the tutor expects to see in an assignment. When closely aligned with learning outcomes, they make clear the connection between the assessment and the stated purpose of the course or module” (Nordberg, 2008, p. 11). For an industry project it is, of course, also important that the client has input into decisions about assessment methods. Willis and Millis (2004) advise lecturers to ask themselves two questions before they finalise their assessment approach: “Should I give students all the same mark [i.e., assess the product only] or a mark based on each person’s individual contribution to the group performance [i.e., assess the process as well]?” and: “If I assess each student’s contribution, how will I know what each person has contributed?” (p. 271).

The simplest and easiest way of assessing group work is, of course, to assess the product only and give the same mark to everyone in the group (Race, 2001). While it is undoubtedly easy to assess the tangible outputs of a group project such as a group report or presentation and to assign a collective group mark, this approach does not address two important issues with extended group work in industry projects. Firstly, a collective group mark does not effectively reflect all the learning objectives that are normally required for this type of group work in that it assesses only the product of the project and not the process. Valid assessment measures must also be devised to evaluate the process aspects of the project, particularly the management of the project, the management of the group, the level of participation and contribution by individual group members, and the personal learning outcomes for each individual (Figure 1).

Secondly, a collective group mark may lead to a perception of unfair assessment for students in that it offers no opportunity for recognition of individual levels of participation and contribution. A collective group mark may not promote effective collaboration in that it does not place value on individual accountability or promotive interaction and may be viewed by students as rewarding freeloading. Nordberg (2008) warns:

The problem with assessing group work is this: the work of the individual is lost in the product. ...Crucially, students regularly question the fairness of the whole process even before they begin to suffer the frequent descent of group work into disagreement and disillusionment. (p. 481)

When a collective group mark is awarded, a group member who has contributed little to the group outputs will receive the same mark as one who has contributed more and “weak students can ride the coat-tails of stronger ones to achieve marks they do not deserve. Worse, perhaps, strong students who contribute more to the group work can be held back from achieving the higher levels of distinction they deserve” (Nordberg, 2008, p. 481). A student in a research study by Clark and Baker (2011) articulated this concern:

I think group assignments and group marks should consider each individual’s levels of participation and contribution. It is unfair to have the same marks for all the group members. People who participate and contribute more should be awarded more marks. (p. 129)

Another student in an earlier study by Clark, Baker, and Chan (2009) expressed a similar sentiment in stronger terms:

Group assignments are indeed unfair. Some people do not do anything, but some people have to do everything. Yet they get the same grade! The lecturers do not care who has done much and who has done little or nothing. (p. 9)

Although the awarding of a collective group mark is often perceived as unfair by students, rarely leads to appropriate student learning behaviour, and frequently results in freeloading (Gibbs, 2010), there is a strong counter argument that giving individual marks in a collaborative activity sends a mixed message to students (University of Queensland Guidelines, n.d.). If lecturers are expecting a group to collaborate, why would they measure and reward individual contributions?

An associated assessment issue articulated by Webb (1993) is the question of whether fairness is achievable in groups if an assessment is intended to measure both quality of the product and the learning process of individuals. If lecturers are assessing the quality of a group product *and* the quality of individual students’ ability to interact, work, and collaborate with others, they might find that “behaviour that is conducive to producing a high-quality group product may not always be conducive to individual learning, and vice versa” (Webb, 1995, p. 5). For instance, the quality of a group’s product does not depend on every group member contributing; it might, in fact, be more efficient and effective for the more able members of the group to complete the task without the involvement of the less able members. Unfortunately, however, if this happens and if all members of the group are given the same mark, both students and lecturers may perceive the result as unfair and complain predictably about ‘bludgers’ and ‘social loafers’. Webb (1995) made the point that penalising these groups for failing to involve all members may in itself be seen as “unreasonable and unfair” (p. 16); it may, therefore, not be possible to measure individual accountability and group productivity in the same assessment without sending students “mixed messages about how they should behave” (p. 27). She suggests that it is essential to clarify the purpose of the assessment and the goal of group work on the assessment, and to make explicit the processes that will help the student achieve that goal; she argues that, if the goal of the group work is high group productivity, then the expected group processes will include division of labour, greater participation by the ablest group members, and a reduced emphasis on individual understanding, and states: “We cannot expect students to concentrate fully on creating a high-quality group product if individual students are being held accountable for their own contributions to group

collaboration” (Webb, 1997, p. 212). If the purpose of the assessment is to measure individual student learning from group collaboration, then it should be designed to encourage processes beneficial for learning and not group productivity.

In student industry projects, however, it is necessary to evaluate group outputs, group collaboration processes, project management processes and individual learning as these are all typically required learning outcomes. It is this need to balance individual and group considerations without giving students mixed messages that makes assessment design for industry projects such a challenging task.

A second issue that must be addressed in assessment design is to incorporate some form of individual accountability and to reduce the opportunity for social loafing. Despite claims that group learning was an effective pedagogical method (Johnson & Johnson, 2009), research consistently documents lecturers and students alike reporting that a major disadvantage of collaborative learning is the propensity for unequal contribution of group members and the consequent unfair allocation of marks (Baker & Clark, 2010; Jassawalla, Sashittal, & Malshe, 2009; Strauss & U, 2007; Williams, Beard, & Rymer, 1991). A large body of research carried out in the workplace has suggested that identifiability and individual accountability (task visibility) is a major factor in reducing social loafing (George, 1992; Harkins & Szymanski, 1989; Williams, Harkins, & Latane, 1981), as is providing reliable performance evaluation mechanisms and clear standards (Harkins & Szymanski, 1989; Szymanski & Harkins, 1987). The conclusion is that if mechanisms masking social loafing are reduced, the behaviour will be moderated. Similar strategies have been found to be successful with student groups. If students are identified by transparent systems of accountability (Oakley, Felder, Brent, & Elhaji, 2004), if the process incorporates peer ratings (Kaufman, Felder, & Fuller, 1999) – particularly multiple peer evaluations (Aggarawal & O'Brien, 2008), and if tasks have reliable performance assessment mechanisms (Michaelson, Fink, & Knight, 1997), the likelihood of social loafing is reduced. Freeloading is discouraged if students are required to keep minutes of meetings and if the lecturer and the client meet regularly with groups to monitor progress (Strauss & U, 2007). These considerations must be factored into the assessment design of student industry projects.

Assessment Methods

The decision to assess both product and process and to moderate individual students' marks on the basis of their contributions to a group project should be carefully considered in advance (Sharp, 2007). While Webb (1997) argued that it is not always possible, or even desirable, to satisfy expectations of “fair” assessment by measuring both productivity and learning processes, researchers such as Felder and Brent (2001) have produced systems that they claim measure both product (a group mark) and process (an individual mark) with integrity. Lejk and Wyvill (1996) noted that, although lecturers may want to give individual grades, prior to their survey the range of methods available to do this effectively had received little attention in the literature; their research surveyed nine different methods of deriving individual grades from group assessments and discussed some of the issues involved. In the following decade, however, the literature has given much more attention to the issue, and researchers such as Millis & Cottell (1998), Willis and Millis (2004), Felder and Brent (2001), and Davis (2009) have produced a range of mechanisms for allocating differential marks to students, usually offered with associated documentation, which are more likely to be perceived as fair by both lecturers and students. Gibbs (2010) commented that lecturers new to assessing group work now have numerous options, including online systems [e.g. WebPA, SPARKPLUS], so do not need to spend time designing their own mechanisms.

Lecturers who decide to assess process might find peer assessment and self-assessment useful methods of identifying individual contributions, particularly if conducted anonymously (Gibbs, 2010). These methods, however, do raise questions of validity and reliability. Falchikov (2003) listed three additional problems with peer or self-assessment in practice: students often dislike the idea, colleagues are often suspicious of the idea, and setting up well-designed systems involves a considerable investment of time.

Although students are in a logical position to judge individual participation and contribution, they may be hesitant to pass judgement on other group members (Johnson & Johnson, 1996). Research suggests that peer and self-assessment can become more meaningful for students when they have input and control over the process themselves by, for example, developing their own criteria for individual grades for contribution to the group project (Willis & Millis, 2004) and that the methods, if implemented carefully, are at least as valid and reliable as lecturer evaluation (Gibbs, 2010). The development of responsible assessments based on measurable criteria, evidence, and objectivity is considered to be crucial if peer and self-assessment systems are to succeed (Willis & Millis, 2004) as is the provision of opportunities for students to develop and practise peer and self-assessment techniques (Kaufman, Felder, & Fuller, 2000). Falchikov and Goldfinch (2000) found that peer evaluations were closer to lecturer evaluations when global judgements based on clear explicit criteria were used rather than evaluation of individual dimensions.

Boud and Falchikov (2005) discussed the vital contribution of self and peer assessment to “sustainable assessment” (assessment which is concerned with preparing the student for future learning and a variety of assessment roles in the contemporary workplace) and stress the importance of preparing students for a lifetime of assessing their own learning. Johnston and Miles (2004) support this view that peer and self-assessment promote independent, reflective, critical learning and are important skills for students to master. While Race (2001) argued strongly for the learning benefits of involving students in their own assessment, he also recognised the risks; he stresses the importance of robust moderation practices as peer assessment can be an easy target for criticism and lecturers must protect themselves.

If peer and self-assessment methods are chosen as a component of industry project assessment, mark allocations and adjustments can be supported by group and individual documentation such as group diaries, meeting minutes, status reports, and individual task allocation and completion documents, journals and reflective work. These can provide the lecturer with more concrete evidence of participation and contribution levels of individual group members and can be matched against the student ratings. It is also assumed that, for an industry project, the lecturer is involved in regular and ongoing monitoring of the groups and will have some understanding of the relative contributions of individual group members. Nordberg (2007) stated that “Diaries, project portfolios recording individual-member contributions and periodic reviews of group progress can all contribute to the tutor's ability to understand which students are working hard and which are hardly working” (p. 3). The use of peer and self-assessment methods can, therefore, address student concerns about recognition of individual levels of participation and contribution.

Young and Henquinet (2000) also recommended the inclusion of parties external to the industry project as an often overlooked assessment source. The client, other project stakeholders and industry professionals can provide valuable input to the assessment process when involved in the design, moderation and marking of both product (reports and presentations by the group) and process elements such as professional behaviour and documentation. Their involvement “establishes a clear connection between the classroom and workplace, and students are often more serious when they know that they will be evaluated by an external expert” (p. 58).

There is no single assessment method for industry projects that can be described as perfect; there are advantages and disadvantages associated with each potential assessment option. What is important is that the chosen assessment design has: content validity, in that it matches the published learning objectives of the course; industry validity, in that it meets industry standards and expectations; and face validity, in that it is easily understood by the students and makes sense to them. In addition, an effective design will address issues of product and process, group and individual accountability, and perceived fairness. The assessment design will normally incorporate a “mix and match” of methods that reflect the complexity of the assessment requirements. A range of assessment techniques that have been identified as appropriate for student group work, along with their advantages and disadvantages, can be found in the Appendix.

An Example of Integrated Assessment Design for Student Industry Projects

An example of an integrated assessment design for a student industry project undertaken in the final year of an undergraduate degree is provided in Figure 2. This was designed for business projects which might involve feasibility studies for new business ventures, market research and analysis, or event planning and management. The design allows for the assessment of both product and process elements; while the majority of the marks are allocated to the project outputs (product), marks are also allocated to the project and group management process and the personal learning outcomes for individual group members (process). The group outputs, in the form of a proposal, a report and presentations are assigned 65% of the total mark which reflects the importance of these outputs for both academic and industry purposes. The client and other industry representatives may be involved in the design, moderation and evaluation of these outputs to ensure industry validity.

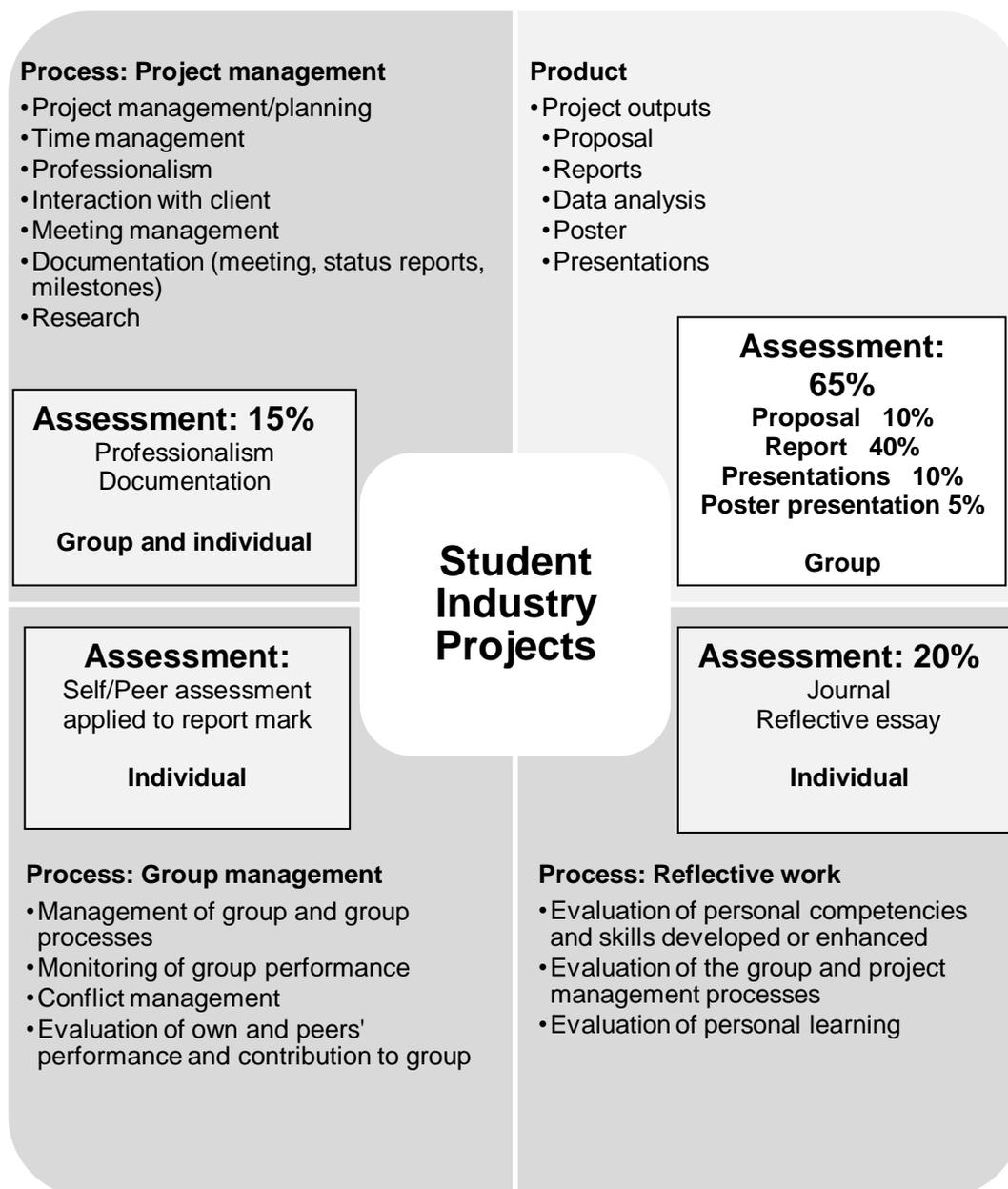


Figure 2: Example of Integrated Assessment Design for Student Industry Project

The self and peer assessment elements of the design allow for adjustment of the report mark to reflect the individual contribution of each group member and signal to students the importance of individual accountability as well as promoting independent, reflective, critical learning and preparing them for future assessment of their own learning. This also addresses issues of fairness in individual participation and contribution.

The assessment of the project and group management processes is based on group documentation (status reports, work breakdown reports and gantt charts, meeting documentation and milestone forms). Again, industry clients may be involved in designing and evaluating the processes and documentation required for the project. Group and individual marks awarded for the standard of professionalism demonstrated by the students are allocated after consultation with the client and other stakeholders in the project and are based on the extent to which the students' outputs, communication, meeting facilitation and primary and secondary research processes meet industry standards.

The reflective work, in the form of a journal and reflective essay, is assigned 20% of the total mark to indicate the importance of the student's recognition and analysis of personal learning and skill development throughout the project process. This completes a design which allows for both academic and industry validity, assesses both product and process, incorporates group and individual elements, encourages accountability, promotes reflective learning and can be perceived as fair by students.

Conclusion

"Assessment is the single most powerful influence on learning in formal courses" argue Boud, Cohen, and Sampson (1999) "and, if not designed well, can easily undermine the positive features of an important strategy in the repertoire of teaching and learning approaches" (p. 413) . In an analysis of research on assessment of group work, Gibbs (2010) came to the same conclusion: group work can achieve the promised benefits "provided that there are associated assessment mechanisms that leverage appropriate student learning behaviour. In the absence of such assessment mechanisms these benefits may well not materialize" (p. 1).

This is significant as tertiary institutions in New Zealand move towards education that is transformational, that is, about what a student is becoming rather than focussing solely on what they know and can do (Wood, Thomas, & Rigby, 2011). Student industry projects are a part of this transformational movement but unless lecturers develop assessment designs that support this aim, many of the benefits of this type of learning may be lost.

References

- Aggarawal, P., & O'Brien, C. L. (2008). Social loafing on group projects: Structural antecedents and effect on student satisfaction. *Journal of Marketing Education*, 30(3), 76-85.
- Aronson, E. (2001, April 8). Jigsaw classroom could help prevent plague of violence. *St. Louis Post - Dispatch*, p. B3.
- Baker, T., & Clark, J. (2010). Cooperative learning: A double edged sword: A cooperative learning model for use with diverse student groups. *Intercultural Education*, 21(3), 257-268.
- Boud, D., & Falchikov, N. (2005). Redesigning assessment for learning beyond higher education. *Research and Development in Higher Education*, 28, 34-41.
- Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment and Evaluation in Higher Education*, 24(4), 413-426.
- Chen, G., Donahue, L. M., & Klimoski, R. J. (2004). Training undergraduates to work in organizational teams. *Academy of Management Learning & Education*, 3(1), 27-40.

- Clark, J., & Baker, T. (2011). It's not fair: Cultural attitudes to social loafing in ethnically diverse groups. *Intercultural Communication Studies*, XX(1), 124-140.
- Clark, J., Baker, T., & Chan, K.W. (2009, June). *Perception is reality: Chinese and New Zealand students' perceptions of cooperative learning*. Paper presented at the International CAFIC Conference, Beijing.
- Davis, W. M. (2009). Groupwork as a form of assessment: Common problems and recommended solutions. *Higher Education*, 58, 563-584.
- Falchikov, N. (2003). Involving students in assessment. *Psychology of Learning and Teaching*, 3(2), 102-108.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70(3), 287-322.
- Felder, R. M., & Brent, R. (2001). Effective strategies for cooperative learning. *Journal of Cooperation & Collaboration in College Teaching*, 10(2), 69-75.
- George, J. M. (1992). Extrinsic and intrinsic origins of perceived social loafing in organizations. *Academy of Management Journal*, 35(1), 191-202.
- Gibbs, G. (2010). The assessment of group work: Lessons from the literature. *Assessment Standards Knowledge Exchange*, Retrieved from www.brookes.ac.uk/asked/documents/Brookes%20groupwork%20Gibbs%20Dec%2009.pdf.
- Harkins, S. G., & Szymanski, K. (1989). Social loafing and group evaluation. *Journal of Personality and Social Psychology*, 56(6), 934-941.
- Jassawalla, A., Sashittal, H., & Malshe, A. (2009). Students' perceptions of social loafing: Its antecedents and consequences in undergraduate business classroom teams. *Academy of Management Learning and Education*, 8(1), 42-54.
- Johnson, D. W., & Johnson R. T. (1989). *Cooperative learning, values, and culturally plural classrooms*. Retrieved from <http://www.cooperation.org/pages/CLandD.html>
- Johnson, D. W., & Johnson, R. T. (1996). *Meaningful and manageable assessment through cooperative learning*. Edina, MN: Interaction.
- Johnson, D. W., & Johnson R. T. (1999). Cooperative learning, values, and culturally plural classrooms. In M. Leicester, C. Modgill, & S. Modgill (Eds.), *Values, the classroom and cultural diversity*. London, England: Cassell PLC.
- Johnson, D. W., & Johnson R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher*, 38(5), 365-379.
- Johnson, D. W., Johnson, R. T., & Smith, K. A. (2007). The state of cooperative learning in postsecondary and professional settings. *Educational Psychology Review*, 19(1), 15-29.
- Johnston, L., & Miles, L. (2004) Assessing contributions to group assignments. *Assessment & Evaluation in Higher Education*, 29(6), 751-768.
- Kaufman, D. B., Felder, R., & Fuller, H. (1999). *Peer ratings in cooperative learning teams*. In *Proceedings of the 1999 Annual ASEE Meeting, ASEE, June 1999*.
- Kaufman, D. B., Felder, R., & Fuller, H. (2000). Accounting for individual effort in cooperative learning teams. *Journal of Engineering Education*, 89(2), 133-140.
- Lejk, M., & Wyvill, M. (1996). A survey of methods of deriving individual grades from group. *Assessment and Evaluation in Higher Education*, 21(3), 267-281.
- Michaelsen, L. K., Fink, L. D., & Knight, A. (1997). Designing effective group activities: Lessons for classroom teaching and faculty development. In D. Dezure (Ed.), *To improve the academy: Resources for faculty, instructional and organizational development*. Stillwater, OK: New Forums.
- Millis, B. J., & Cottell, P. G. (1998). *Cooperative learning for higher education faculty*. Phoenix, AZ: The Oryx Press.
- Nordberg, D. (2007). Fairness in assessing group projects: A conceptual framework for higher education. *International Journal of Quality and Standards*, 1(2), article 15.
- Nordberg, D. (2008). Group projects: more learning? Less fair? A conundrum in assessing postgraduate business education. *Assessment & Evaluation in Higher Education*, 33(5), 481-492.

- Oakley, B., Felder, R. M., Brent, R., & Elhadj, I. (2004). Turning student groups into effective teams. *Journal of Student Centred Learning*, 2(1), 9-34.
- Race, P. (2001). Assessment: A briefing on self, peer and group assessment. *LTSN Generic Centre Assessment Series Number 9*, York: LTSN Generic Centre.
- Sharp, S. (2007). Deriving individual student marks from a tuto's assessment of group work. *Assessment & Evaluation in Higher Education*, 31(3), 329-343.
- Slavin, R. E. (2001). Cooperative learning and intergroup relations. In J. Banks & C. McGee Banks (Eds.), *Handbook of research on multicultural education* (pp. 628-634). San Francisco, CA: Jossey-Bass.
- Strauss, P., & U, A. (2007). Group assessments: Dilemmas facing lecturers in multicultural tertiary classrooms. *Higher Education Research and Development (HERD) Journal*, 26(2), 147-161.
- Szymanski, K., & Harkins, S. G. (1987). Social loafing and self evaluation with a social standard. *Journal of Personality and Social Psychology*, 53(5), 891-897.
- University of Queensland guidelines. (n.d.). *Group assessment: Assessment of students on group based tasks-issues and options*. Retrieved from <http://www.tedi.uq.edu.au/assess/Assessment/groupwork.html>
- Webb, N. (1993). *Group collaboration in assessment: Competing objectives, processes and outcomes*. Washington, DC: National Center for Research on Evaluation, Standards and Student Testing.
- Webb, N. (1995). Group collaboration in assessment: Multiple objectives, processes and outcomes. *EEPA*, 17(2), 239-261.
- Webb, N. (1997). Assessing students in small collaborative groups. *Theory into Practice*, 36(4), 205-212.
- Williams, D. L., Beard, J. D., & Rymer, J. (1991). Team projects achieving their full potential. *Journal of Marketing Education*, 13, 45-53.
- Williams, K., Harkins, S., & Latane, B. (1981). Identifiability as a deterrent to social loafing: Two cheering experiments. *Journal of Personality and Social Psychology*, 40, 303-311.
- Willis, D., & Millis, B. J. (2004). An international perspective on assessing group projects. *To Improve the Academy*, 22, 268-286.
- Wood, L. N., Thomas, T., & Rigby, B. (2011). Assessment and standards for graduate outcomes. *Asian Social Science*, 7(4),12-18.
- Young, C. B., & Henquinet, J. A. (2000). A conceptual framework for designing group projects. *Journal of Education for Business*, 76(1), 56-60.

Appendix: Assessment Methods for Student Industry Projects

ASSESSMENT OPTION	DESCRIPTION	ADVANTAGES	DISADVANTAGES	SCOPE
Assessment of product only.	The same mark for everyone in the group.	<ul style="list-style-type: none"> • Encourages collaboration and positive interdependence as groups sink or swim together. • Straightforward method, easily understood. • Cuts down marking. 	<ul style="list-style-type: none"> • Individual contributions are not reflected in the marks. • No recognition of individual excellence or poor individual performance. • Stronger students may be unfairly disadvantaged by weaker students. • May not motivate all students; can encourage social loafing. • Method most often perceived as unfair by students and lecturers. 	Product
Assessment of product with lecturer adjustment.	Assessment of product. The same mark for everyone in the group. Direct evaluation by lecturer to modify the group mark for individual group members. <ul style="list-style-type: none"> • oral interviews • meeting minutes • observation • group reports 	<ul style="list-style-type: none"> • Oral interviews and observation can be an effective way of getting information on individual participation. • Enables the lecturer to give each student specific feedback. • Encourages students to reflect on their own performance. 	<ul style="list-style-type: none"> • Very time consuming. • Class size might make it infeasible. • Information obtained may be subjective or inaccurate. • Reliability and validity cannot be guaranteed. 	Product and process
Group average mark	Individual submissions (allocated tasks or individual reports) are marked individually. The group members each receive the average of the individual marks.	<ul style="list-style-type: none"> • May provide motivation for students to focus on both individual and group work. 	<ul style="list-style-type: none"> • May be perceived as unfair by students • Stronger students may be unfairly disadvantaged by weaker students. 	Product

ASSESSMENT OPTION	DESCRIPTION	ADVANTAGES	DISADVANTAGES	SCOPE
Individual marks for allocated project tasks.	Project tasks are divided up and allocated to individuals; the parts are marked separately.	<ul style="list-style-type: none"> • Ensures individual motivation and participation. • Rewards good performance and penalises poor performance. • May be perceived as fairer than a shared mark. 	<ul style="list-style-type: none"> • Difficult to find tasks that are exactly equal in size/complexity. • Does not encourage the group process or collaboration. • Dependencies between tasks may slow progress of some students. 	Product
Individual marks for individual reports	Each student writes and submits an individual report based on the group's work.	<ul style="list-style-type: none"> • Ensures individual effort • Perceived as fair by students 	<ul style="list-style-type: none"> • Precise manner in which individual reports should differ often very unclear to students • Likelihood of plagiarism increased 	Product
Individual marks based on a subsequent examination/ test	Individual marks based on an examination/ test held after group project is finished. The mark is added to the group mark.	<ul style="list-style-type: none"> • May increase motivation to learn from the group project and to learn from other members of the group. • May encourage the group process and collaboration. • May be perceived as fairer than a shared mark. • Identifies slackers. • Allows deserving students the opportunity to shine. 	<ul style="list-style-type: none"> • May diminish the importance of group work. • Gives additional work to lecturers. • May unduly reward students who are good at written examinations and tests but may have underperformed in their group. 	Product

ASSESSMENT OPTION	DESCRIPTION	ADVANTAGES	DISADVANTAGES	SCOPE
Group mark for product differentiated by group peer assessment	<p>One mark for the group outputs. Subsequent adjustment to acknowledge individual contribution based on peer assessment:</p> <p>a) Students redistribute a pool of marks; the group has to explain its decisions. Team members score their relative contribution out of 100. If all worked equally hard they are all allocated 100% of the mark for the assessment. If one team member contributes only half as much, that team member is given 50% of the marks.</p>	<ul style="list-style-type: none"> • Shares the assessment responsibility; group members are usually in a better position to judge relative contribution to the group. • Easy to implement. • May motivate students to contribute more. • Transferable negotiation and appraisal skills are developed. • Has the potential to reward good performance and penalise poor performance. • May be perceived as fairer than a shared mark. • Helps reduce social loafing. • Puts value on individual contributions. • Gives ownership of decisions to the group. • May be perceived as fairer than a shared mark. 	<ul style="list-style-type: none"> • Subjective evaluation may occur. • May lead to conflict that students are unable to manage. • Students may not have the required negotiation and conflict solving skills. • Can be intimidating to some students. • Requires training and practice for students. • Can result in everyone just agreeing to have the same mark to avoid unpleasantness. • Must be monitored carefully by the lecturer. • May allow the group to discriminate against individuals. • Resistance by some students. • Public peer assessment has been shown to be less effective than anonymous peer assessment. 	Product and process

ASSESSMENT OPTION	DESCRIPTION	ADVANTAGES	DISADVANTAGES	SCOPE
Group mark for product differentiated by individual peer assessments	<p>One mark for the group outputs. Subsequent adjustment to acknowledge individual contribution based on peer assessment:</p> <p>b) A peer assessment form is filled in by all students and the lecturer adjusts the marks accordingly. Can be done anonymously.</p>	<ul style="list-style-type: none"> • Encourages a sense of involvement, responsibility and accountability. • Helps students develop appraisal skills. • Group members are usually in a good position to judge relative contribution to the group. • Has the potential to reward good performance and penalise poor performance. • Easy to implement. • Can motivate students to contribute more. • May be perceived as fairer than a shared mark. • Reduces social loafing. • Puts value on individual contributions. • Gives ownership of decisions to individuals in the group. 	<ul style="list-style-type: none"> • Students need training and practice to carry out the process effectively. • Lecturer moderation is time consuming. • Open to subjective evaluations; can be perceived as unfair and based on students' popularity or unpopularity. • Some students are reluctant to mark down peers. • Group members may agree on an equal mark to avoid conflict. • Can be seen as inhibiting co-operation. • Has to be monitored carefully. • Some students don't want this responsibility. • May be difficult to translate into marks. 	Product and process
Group mark for product plus individual contribution mark allocated by group	<p>Contribution marks are added to the final product mark.</p> <p>A mark is given by the lecturer for the product and group members are asked to peer assess an additional mark for contribution for each group member.</p>	<ul style="list-style-type: none"> • May be perceived as fair. • Gives the message that process is important. • Doesn't put complete responsibility on the group members. • Final mark can weight group performance more heavily to encourage collective effort. • Helps students develop appraisal skills. 	<ul style="list-style-type: none"> • Students need training and practice to carry out the process effectively. • Some students are reluctant to mark down peers. • Group may agree on the same mark for all to avoid conflict. • Some students don't want this responsibility. • Some students perceive peer assessment as being too subjective. (This concern is not backed up by research.) 	Product and process

ASSESSMENT OPTION	DESCRIPTION	ADVANTAGES	DISADVANTAGES	SCOPE
Group mark for product plus self-assessment	<p>Self-assessment: One mark for the group outputs.</p> <p>Students assess their own contribution to the work of the group and the lecturer adjusts the marks accordingly. Usually carried out in conjunction with peer assessment.</p>	<ul style="list-style-type: none"> • Encourages personal responsibility and self appraisal. • May develop self-awareness. • May develop better understanding of learning outcomes. 	<ul style="list-style-type: none"> • Students need training and practice to carry out the process effectively. • Students can tend to make judgements based on what they meant to do rather than what they actually achieved. • It may be difficult to define common standards for students to use for evaluation. • Some lecturers and students believe it is not reliable or valid (though research does not back this up.) 	Product and process
Addition of assessed reflective work	<p>Reflective work to identify individual learning outcomes:</p> <p>The addition of assessed reflective work (such as essays, journals, diaries, portfolios) to other assessment methods.</p>	<ul style="list-style-type: none"> • Encourages individual appraisal of personal development of required knowledge and skills. • Recognition of transferable skills. • Encourages personal responsibility. • May develop self-awareness. • May develop better understanding of learning outcomes. 	<ul style="list-style-type: none"> • Students need training and practice to carry out the process. 	Process

Adapted from:

Carnegie Mellon. (n.d.). Grading methods for group work. Retrieved from <http://www.cmu.edu/teaching/assessment/howto/assesslearning/groupWorkGradingMethods.html>

Winchester-Seeto, T. (2002, April). *Assessment of collaborative work - collaboration versus assessment*. Invited paper presented at the Annual Uniserve Science Symposium, The University of Sydney.

Popular Culture, Overall Teacher Judgments and National Standards

John Dickie and Mary Jane Shuker
Victoria University of Wellington

Email: john.dickie@vuw.ac.nz; mary-jane.shuker@vuw.ac.nz

Abstract

With the introduction of national standards in New Zealand schools for years 1 to 8 educators are required to make overall teacher judgments of children's reading, writing and mathematics achievement and progress. This paper explores how teachers of years 1 to 3 children capitalised on their interest in, and knowledge of, popular culture in school literacy activities and to what extent they used this in forming their judgments of children's literacy skills and knowledge. While Ministry of Education guidelines describe literacy as socio-cultural practice, they include lists of skills and knowledge for each year level, suggesting a more skills-driven teacher-centred practice. Teachers can reconcile this apparent difference by capitalising on children's own expertise in popular culture. We ask the question, to what extent do primary teachers incorporate children's interest and expertise in popular culture as they assess literacy? Our investigation gathered data through a comprehensive survey followed by in-depth case studies, which involved interviews with teachers and children, observations and document analysis. A majority of teachers in our study reported that they incorporated aspects of popular culture as they taught and assessed literacy in their classroom programmes.

Introduction

In early 2010, the New Zealand government introduced national standards in reading, writing and mathematics for the first eight years of compulsory schooling. A policy paper on assessment was released to explain that:

Unlike standards-based assessment in other countries, our standards do not rely on national testing. Instead there is an emphasis on teacher professional judgments, assessment for learning principles and practice, and the importance of information sharing to support student learning. This is a novel approach when compared with other jurisdictions. (Ministry of Education, 2010a, p. 3)

When exploring overall teacher judgments (OTJs) and national standards in relation to children's literacy in primary schools it is necessary first to examine the theoretical views in the official documents that drive the teaching and assessment. The Ministry of Education's requirements for primary school literacy are promulgated through *The New Zealand Curriculum* (Ministry of Education, 2007) and its supporting materials, which have increasingly shown an interpretation of literacy as social and cultural practice. Over the past two to three decades there has been a shift from what Gee (1996) described as a traditional view of literacy based on psychological approaches, which tended to focus more on the technical skills of literacy, to an anthropological approach which explores literacy as social practice. Gee writes that the traditional view of literacy treats it as an asocial cognitive skill with little or nothing to do with human relationships. He explains the shift in interpretation of literacy from a largely psychological interpretation to one of social practice: "The traditional view of literacy has defined it in rather simple terms: literacy is the ability to read and (sometimes) to write" (p. 39). This involves decoding and encoding and interpretation of the text. Gee writes that for traditionalists the interpretation is largely a psychological matter: "If readers know the language, can decode writing, and have the requisite background 'facts' ... they can construct the 'right' interpretation in their heads. And this 'right' interpretation is roughly the same for all competent readers" (p. 39). Current conceptualisations of socio-cultural theory draw on the work of Vygotsky (1978)

who explains how the social, cultural and historical background and experiences help shape children's developing cognition.

Successive Ministry of Education literacy handbooks for teachers have demonstrated an increasing emphasis on the need to make effective links to children's cultural backgrounds; this seems appropriate considering the growing diversity of children in New Zealand classrooms. This approach is consistent with the work of Street (1984) whose ideological model of literacy concentrates on the specific social practices of reading and writing, recognising the culturally embedded nature of these. Traditionally, reading and writing have often been understood as discrete cognitive skills (Gee, 1996) which Street (1984) describes as the autonomous view of literacy. Street explains this interpretation of literacy as a neutral technology, in other words a set of skills that are culture free, and which can be detached from specific social contexts. In New Zealand, as in other countries, the debate about how reading and writing are best taught has focused largely on skills but this narrow interpretation has been challenged by a broader and more complex view of literacy as social practice (see, for example, Makin, Jones Diaz, & McLachlan, 2007).

This interpretation of literacy supersedes traditional views of literacy, which emphasised the acquisition of skills, with a socio-cultural approach that looks at literacy from the point of view of social practice. In the 1970s and 1980s a group of scholars began to question the traditional views of literacy and developed a new interdisciplinary field of study referred to as the New Literacy Studies (Gee, 1996). This approach, which was strongly influenced by the early work of Heath (1983) and Street (1984), focuses on the social, cultural, and political contexts of the users rather than on the technical skills of reading and writing. Literacy is seen as "multiple", and reading and writing are differently influenced and transformed in various socio-cultural practices. If we are to take an approach to literacy influenced by the New Literacy Studies and socio-cultural theory we can consider how teachers may draw upon children's out-of-school expertise and interests to make effective connections to the requirements of school literacy.

According to Lankshear and Knobel (2003), the New Literacy Studies use a socio-cultural approach to understanding and researching literacy, offering a new paradigm for exploring literacy as opposed to one based on psychology, which could be described as a psycholinguistic or a technicist paradigm. Lankshear and Knobel point out that the socio-cultural approach to literacy rejects the idea that textual practices are largely or solely processes which go on in the head or are communication from one head to another by means of graphic symbols. A socio-cultural perspective focuses on social practices. The New Literacy Studies argue that children are socialised into particular practices of literacy, and at the same time socialised into discourses that position them ideologically within the larger social and cultural environment (Razfar & Gutierrez, 2003). In agreement with Robbins (2005), when adopting a socio-cultural approach it is necessary to look at context, relationships, culture and the activities in which the participants are involved as well as the tools and artefacts they use.

Consistent with this interpretation of literacy acquisition is the strategy of incorporation (McNaughton, 2002). This enables transfer of learning for children between out-of-school and school sites. McNaughton argues that for children from culturally and linguistically diverse backgrounds who are beginning school, the "meeting of minds" between teacher and learner is vitally important for effective literacy teaching (p.14). This can be achieved through "enhancing the multitude of sites that learners and teachers have available to them in which to make connections" (p. 26). McNaughton states: "transfer of learning occurs as a consequence of this incorporation – bridges between the familiar and the unfamiliar can be made by both the learner and by the teacher" (p. 27). However, incorporating out-of-school knowledge and expertise may not match the curriculum or pedagogy of the school. This can present a challenge for teachers in their classroom practices.

Literacy Assessment in Primary Schools

While accurate encoding and decoding are essential skills, the main purposes of writing and reading are concerned with meaning. The *Effective Literacy Practice* handbooks (Ministry of Education, 2003, 2006) take a social practice interpretation of literacy and advise teachers to become better informed of children's out-of-school literacies. Knowledge of these, it is argued, can assist teachers to make connections between children's out-of-school practices and school literacy, and thus improve learning for these children.

Schools are now required to provide more detailed information on student achievement to the Ministry of Education. In December 2008, the newly elected National coalition government made changes to the Education Act, updating the National Education Guidelines to schools. This was to include measurement of student achievement against national standards in reading, writing and numeracy. The National Standards (Ministry of Education, 2009) were to be implemented in schools from 2010. The introduction has been controversial with the primary teachers' union campaigning against them and some school boards of trustees refusing to implement them. Criticisms were that the standards were not trialled before implementation, that teachers within schools and between schools would interpret them differently, and that the government would use the data to establish league tables to compare school results. Each standard, written for each of the first eight years of compulsory schooling, is a short and relatively simple statement. Teachers are required to report on student achievement in relation to the standards for their year group, as being above, at, below or well below, the standard. The standards for the first year of school are stated below.

The reading standard

After one year at school, students will read, respond to, and think critically about fiction and non-fiction texts at the Green level of Ready to Read. (Ministry of Education, 2009, p. 20)

The writing standard

After one year at school, students will create texts as they learn in a range of contexts across the New Zealand Curriculum within level 1. Students will use their writing to think about, record, and communicate experiences, ideas, and information to meet specific learning purposes across the curriculum. (Ministry of Education, 2009, p. 21)

Each reading standard is supported by a list of characteristics of texts at this level and exemplars of suitable texts. Each writing standard is supported by a list of characteristics of students' writing as well as exemplars of writing. Considerably more detail of the required knowledge and skills is found in *The Literacy Learning Progressions* (Ministry of Education, 2010b), which were developed and trialled before the standards were introduced. The progressions, which guide teachers of students in years 1 to 10, are consistent with the *Effective Literacy Practice* handbooks (Ministry of Education, 2003, 2006) in stating that "Literacy is a socio-cultural practice" and that teachers actively seek opportunities to build on the students' own experiences of culture, language and identity that they bring to the classroom (Ministry of Education, 2010b, p. 7). Although the reading and writing standards document (Ministry of Education, 2009) draws its theoretical base from the Ministry of Education handbooks, it places much less emphasis on linking to children's own backgrounds and interests. It does state that "when teachers are clear about the reading and writing demands of each curriculum area and students' lived experiences (in relation to culture, language, and identity) they can deliberately integrate the teaching of literacy with the teaching of curriculum content in appropriate ways" (p. 5). The document states that teachers should select tasks and texts that reflect students' own

experiences. However, the progressions themselves are lists of knowledge and skills and, if a socio-cultural interpretation of learning is followed, the challenge for teachers is to make learning the required knowledge and skills relevant for the learners.

A position paper on assessment (Ministry of Education, 2010a), written to inform and direct policy, states that the implementation of the national standards “relies on professional qualitative judgments made by teachers on the basis of what they have learned about a student’s achievement and progress across a range of assessment information (both formal and informal)” (p. 14). It makes the important point “*There is no national test. This is a novel approach when compared with approaches elsewhere in the world*” [italics added] (p. 14). The following figure illustrates that no one tool, activity or observation will be sufficient to provide the necessary information.

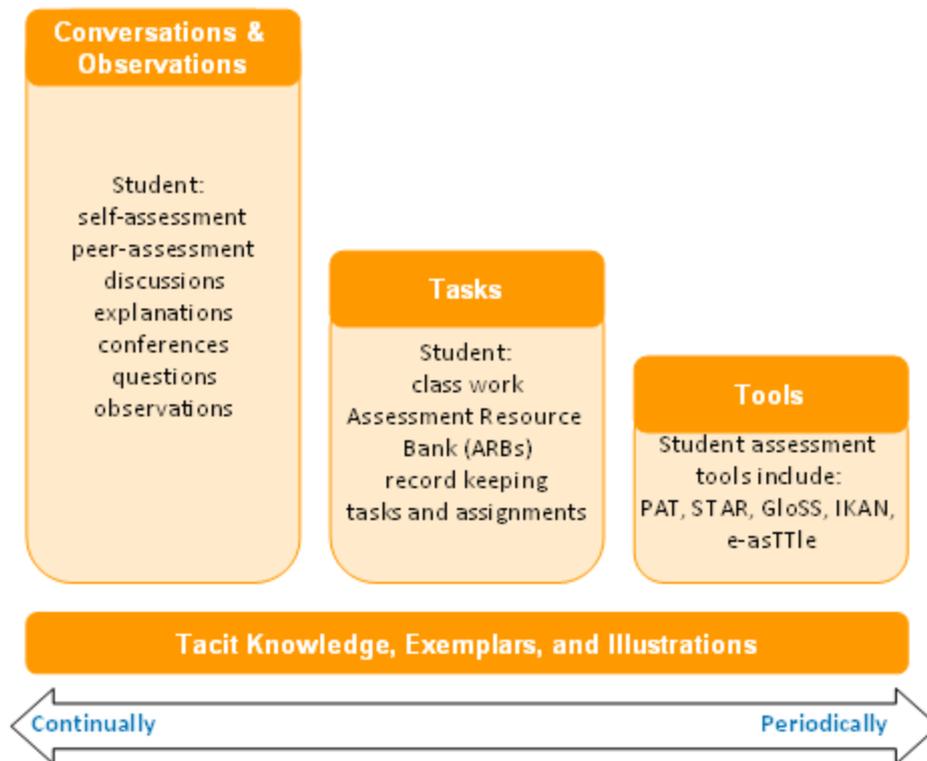


Figure 1: Assessment Tools and Activities
(Ministry of Education, 2010a, p. 35)

The assessment tools and activities outlined here include continual conversations and observations (discussions and conferences with students, student self and peer-assessment and questions) as well as students’ class work. In addition, “At specific points in the school year, tools will be used to provide more formal quantitative assessment” (Ministry of Education, 2010a, p. 36).

These tools and activities can provide opportunities for teachers to make connections (McNaughton, 2002) to students’ own backgrounds and interests as they assess their reading and writing. More formal assessment tools, some of which are norm-referenced, will be used less frequently. Teachers are required to make “overall teacher judgments” by analysing data from these sources and deciding on the “best fit” of how each student’s performance meets the relevant standard (Ministry of Education, 2009, p. 13). Teachers are also required to “specifically consider how well each student is using reading and writing as interactive tools to enable them to learn in all curriculum areas” (p. 13).

As already stated, the Ministry of Education’s requirements for primary school literacy are promulgated through *The New Zealand Curriculum* (Ministry of Education, 2007) and its

supporting materials which have increasingly shown an interpretation of literacy as social and cultural practice. This paper considers the “official” school curriculum through a socio-cultural lens. Within a primary school setting teachers can take into account their children's “lived experiences” (Ministry of Education, 2009, p. 7). However, achieving the links to out-of-school practices and interests is made more difficult when education is increasingly driven by assessment if this is concerned with school literacy practices only.

While it is possible that the imposition of standards in reading and writing may narrow teaching of the curriculum in New Zealand, primary teachers still retain considerable freedom in how they will teach it. For example, there are no basal readers or prescribed programmes. If teachers follow a socio-cultural interpretation of children's learning and wish to make effective connections through incorporation, an obvious way to achieve this is by drawing on children's own interests and expertise in popular culture.

Popular Culture

The perspective of popular culture used here draws on the work of Dyson (1997, 2001, 2003), Luke (1997), Marsh (2005, 2006, 2009), Marsh and Millard (2000) and Morrell (2002). Literacy associated with popular culture involves not only distinctive technical skills but also social skills. Marsh and Millard (2000), for example, describe children's popular culture as overlapping that of adults in the broad fields such as music, sport, computer software, books, magazines, and film. They point out that children's popular culture also includes many diverse artifacts such as toys, games, comics, cards, clothing, sports accessories, and food and drink, and these cultural forms are constantly emerging and disappearing. Luke (1997) depicts children's popular culture as television, video games, comics, and the toys associated with television and movies. She maintains that from these texts of popular culture children develop from infancy their understandings of such things as good versus evil, heroes, gender, race and social power in society. Just as the concept of culture can be interpreted widely, so can children's popular culture.

Knowledge of children's out-of-school uses of literacy in social and cultural settings including popular culture enables teachers to incorporate children's expertise into school and make effective connections for children from diverse backgrounds (Dickie, 2008). Teachers need to be aware of the consonant and conflicting values about popular culture that students face between sites such as family and school. Where there are contradictions between sites, this may provide useful links for teachers in presenting a challenge to adapt the school curriculum in order to make meaningful connections to school (Dickie, 2011).

This study was part of a larger research project that we undertook to examine how teachers can make effective links to children's out-of-school knowledge by incorporating aspects of popular culture into early childhood and school literacy programmes for children aged 12 months to 8 years. This paper focuses on the data from primary teachers and children aged 5 to 8.

Methodology

Participants

Data were collected in two phases. Phase one consisted of a comprehensive postal survey sent to teachers of children in years 1 to 3 in 154 primary schools in the greater Wellington region in New Zealand. The questionnaire was piloted by two primary teachers who responded well to the questionnaire. However, minor changes were made as a consequence of this consultation. There was no follow-up to the mail-out questionnaire but five prizes were given out to maximise the return rate.

Forty-two teachers returned the survey, from 30 schools, a response rate of 27 percent. They were asked a range of questions concerning their attitudes, beliefs and experiences regarding children's use of popular culture and to what extent they incorporate this

expertise in their classroom programmes. Questions also focused on demographic data such as characteristics of their school, the children and families who attend, and the ethnic groups they belonged to. The questionnaire was a mixture of multiple choice and open-ended questions to enable ease of completion. In addition, the questionnaire invited respondents and the children in their classrooms to be involved in the second phase of the study. Twenty-three teachers responded to this invitation.

Questionnaire data were analysed to identify participants for the second phase of the study. We selected teachers who provided multiple and detailed examples in the open-ended questions of children using characters, plots or ideas from popular culture in activities at their school. Four female teachers, including two from one school, agreed to take part in case studies to explore in more depth the links between popular culture and literacy. This included individual semi-structured interviews where we asked how they incorporated aspects of children’s popular culture into literacy learning and their views on how children do this by themselves. The interviews were done in person, completed within one hour, and held outside of schools hours. Each interview was audio recorded with the permission of the interviewees and transcribed. The teachers were given their transcript to check for accuracy and make any changes. Only one teacher did so.

Group semi-structured interviews were also carried out by the researchers with a total of 31 children (see Table 1), selected by the teachers based on their use of popular culture in the literacy curriculum. It was explained to the children that we wished to ask questions about their ideas on how they make literacy links to popular culture. The group interviews took place in a designated room in each school and were audio recorded, then transcribed. The interviews allowed us to explore in depth teachers’ and children’s use of popular culture in their literacy curriculum.

Table 1: Age and Gender of Children Interviewed

Age	Male	Female
5 years	2	
6 years	5	4
7 years	8	11
8 years	1	
	Total 16	Total 15

Observations of children writing about popular culture in their classrooms were another source of data. This allowed us to examine how children were using characters, plots or ideas from popular culture in literacy activities. We talked with the children about their work and the specific connections they were making to popular culture. This meant that we did not have to rely on another person’s interpretation of what was happening in the classroom.

Documentation was another source of data collection used in this study. As Yin (2009) notes, documentation is important to use to corroborate and augment evidence drawn from other sources. We photocopied examples from children’s writing books that incorporated aspects of popular culture. Photos were also taken of popular cultural examples in each of the classrooms such as school lunch boxes, school bags, board games and so on based on television and movie characters. These were used to provide an additional source of information for the triangulation of data.

Procedures and approach to analysis

Socio-cultural theory was used for the design of the survey as well as the analysis of data. It informed the design of the questions which sought information from teachers about links to popular culture in classroom literacy. Statistical analysis of the data from the survey was undertaken by a research assistant using Statistical Package for Social Sciences (SPSS).

Qualitative data including the open-ended questions from the survey, interviews, observations and documents were analysed by a process of category construction, according to the question responses. Common themes were sorted into categories from which deeper analysis using a socio-cultural lens revealed sub categories. We then compared these categories, identifying key patterns as well as relationships among them. By analysing how the categories linked together we sought to develop themes of how teachers of years 1 to 3 children capitalised on their interest in, and knowledge of, popular culture in school literacy activities and the extent to which they used this in forming their judgments of children's literacy skills and knowledge.

Ethics

This study received approval from Victoria University of Wellington and adhered to all the usual ethical requirements for informed consent, voluntary participation and confidentiality. Teachers' and children's responses have been anonymised; real names and other identifying information have not been used in the reporting of the findings of this study.

Limitations of the Data

The return rate of the surveys was relatively low at 27 percent; however, we did ask the teachers to complete the surveys on a voluntary basis during a busy time of the school year.

It is important to point out that our study was small – a survey was carried out in the Wellington region and only three case studies were undertaken. Therefore, the results of this study cannot be generalised to schools in New Zealand and can make no claims to how teachers incorporate aspects of popular culture as they teach and assess literacy in their classroom programmes. Another limitation was that three of the four teachers selected for the case studies were from schools located in high socio-economic areas; thus the data do not represent a cross-section of New Zealand schools.

Results

This paper reflects on the findings from the survey and the interviews with three of the teachers and children aged 5 to 8 about use of popular culture in the literacy curriculum.

The majority of the teachers who completed the survey considered that it was appropriate and useful to incorporate popular culture into classroom programmes and they acknowledged children's expertise and interest in this area. In contrast, a small number viewed popular culture negatively and perhaps as something that children should be protected from. The majority of examples from teachers were boys' preferences and some referred to children acting out roles (mainly boys) at interval and lunchtime.

In the survey, teachers were asked to indicate to what extent they agreed or disagreed with a series of statements by marking the response most clearly matching their teaching experiences and/or how they felt. Their choices indicated whether they strongly disagreed, disagreed, somewhat disagreed, somewhat agreed, agreed or strongly agreed with each statement.

The following two statements, which were asked in the survey, were worded to elicit a critical response from teachers about the use of popular culture in their classrooms. A majority of these teachers were positive about incorporating aspects of popular culture. In response to the statement *Reading and writing standards are likely to discourage teachers from incorporating popular culture into class programmes* there were 18 teachers who agreed to some extent while 24 disagreed to some extent. In response to the statement *Reading and writing standards won't interfere with teachers' ability to link to popular culture* there were 28 who agreed to some extent while 12 disagreed to some extent. This range of views included five who strongly agreed and three who strongly disagreed.

Making Overall Teacher Judgments about Reading and Writing

Teachers were asked to comment on the question Do you consider that assessment of children's reading and writing that make links to popular culture can contribute to your overall teacher judgment of children's standards of reading and writing? Of the 42 teachers, eight disagreed with one stating "I believe popular culture is without depth" while another wrote:

Children, especially boys, who write largely on their passion about computer games, Wii etc. find it hard to look beyond the surface features and their writing doesn't have the depth seen in other children's writing. These children often have a diet of these experiences instead of a balance of literature, books and pop culture.

Six responded with "no" but did not give any explanation for their view.

Thirty teachers, however, agreed with some expressing their views strongly. Among the teachers who agreed there were several common responses. Links to popular culture were described as motivating children (7 responses) and engaging them in learning. One answered, "Yes definitely – need to go with what interests and motivates students". Others mentioned enthusiasm and the "importance of hooking into the students' world". Arguments for making links to popular culture included that the subject matter was not as relevant as the textual features of what was being assessed. Another argument was that linking to popular culture could contribute to overall teacher judgments as "it gives a reflection of what a child can write unaided which is a better indication of their independent writing than teacher-led writing tasks". One teacher agreed but cautioned that teachers needed to ensure writing had the necessary features and texts for reading needed to be of suitable quality. Another teacher's justification was "Assessment of reading and writing in any context including popular culture should be part of a teacher's OTJ. We need a broad picture of children's reading and writing in a variety of situations".

Teachers were also asked to respond to the statement *Making links to popular culture is appropriate in school programmes in what is taught and what is assessed*. Thirty-two teachers out of 42, a clear majority, agreed to some extent with this statement while eight disagreed to some extent. There was a similar result for the statement *Incorporating aspects of popular culture helps children to make sense of literacy in school literacy programme*, with 31 teachers agreeing to some extent and the remainder disagreeing to some extent. Of these, 5 teachers strongly agreed while only 1 strongly disagreed.

It seemed, though, that fewer teachers thought it was part of their school programme. When they responded to the statement *Incorporation of literacy links to popular culture is part of our school's literacy programme*, there were 25 teachers who agreed to some extent while 17 disagreed to some extent.

In the survey teachers were invited to give examples of children using characters, plots or ideas from popular culture in activities at school. The most popular characters were from television and movies and of these Ben 10 was mentioned most often (12 times). Ben is a young American boy (cartoon character) who uses a device like a wristwatch (an omnitrix) that enables him to turn into various extra-terrestrial forms in order to battle the aliens who are trying to take over the world. Children who watch this are familiar with his human helpers and the many alien forms Ben transforms into and their special characteristics.

Various superheroes were mentioned 14 times. These included Spiderman, Superman, Batman, Power Rangers and Zac Power. Harry Potter was mentioned six times. Teachers also listed Indiana Jones, Star Wars, Spongebob Squarepants, Wrestling, transformers, vampire programmes, as well as Dora the Explorer, Barbie, Princesses, fairies and ponies.

Popular Culture Characters in Children's Writing

The survey revealed that the majority of school activities associated with these popular culture characters involved children's writing (17 examples) and what teachers described as playground games or children acting out roles (15 examples).

When asked how topics are selected for children's writing, 25 of the 42 teachers indicated that children had some degree of choice, though this lessened as children moved into more structured programmes in classes beyond year 3. Play seemed to be an integrated component of children's use and understanding about popular culture. While games are not in themselves literacy activities they were at times associated with children's topics for writing as well as with drawing symbols such as icons from their favourite programmes. One teacher's example illustrates this: "Year 3 boys drawing with spiky hair from Ben 10 characters! Ben 10 is acted on in imaginary play at lunch time. Story writing involving aliens is popular at the present time."

The Phase 2 interviews with children confirmed their interest in Ben 10. In School 1 the girls and boys enthusiastically read out their brief stories about the Ben 10 characters. The girls enjoyed writing about Ben's cousin Gwen who plays a key role in the fight against the aliens. Their teacher commented that all the children in her class wore Ben 10 clothing.

Although it is not directly concerned with literacy, teachers in the survey frequently described how children acted out roles in imaginary play during interval and lunchtime. These popular culture links were to television and movie characters. Examples included princesses, fairies and ponies for girls and transformers and superheroes for boys. Some of this dramatic play was later used for plots in writing (see Table 2).

Table 2: Characters Named by Teachers in Children's Writing

Television Shows	Movies	Books/ Magazines	Video Games	Toys	Sports Figures	Pop Stars
Ben 10	Shrek	Harry Potter	Aliens	Bratz Dolls	All Blacks	Michael Jackson
Sesame Street (Bert and Ernie)	Lord of the Rings	Fairies	Vampires	Action Men	Big time Wrestling Characters	Justin Bieber
	Harry Potter	Zac Powers	Digimon	Transformers		Ludacris
Cartoon characters	Transformers (2)	Ponies	Pokemon		Sports teams/players	
Star Wars	Current movie characters	Star Wars	Coyote Bandicoot			
Spiderman	Star Wars					
	Spiderman					

One of the teachers interviewed from School 2, Greta, explained how her children's interest in popular culture helped her in making an overall teacher judgment of their writing. She stated:

As one of our writing topics last year we wrote character and so we started out with a familiar character for everyone which was Greedy Cat, the Joy Cowley character and then we moved on to the children independently writing character descriptions so we taught features of a character description then they would choose a character and I remember one of my little boys writing about Batman. He chose to write about Batman and he used all the Batman language like Kapow and Boom and I used that as his writing assessment.

Greta clarified this sample was assessed against a writing matrix that she and her colleagues had developed, which was in line with the literacy learning progressions. She said that the national standards were taken into account in creating the matrix. When Greta was asked what was the advantage of allowing the child to write about his choice, Batman, when assessing him against national standards, she responded:

I think they will be much more motivated especially with boys in the year 1 and 2 classroom. I have a small group who don't feel very happy towards writing. They don't enjoy it very much because it's hard and if they are writing about something that they are really interested in like Batman it's actually much easier for them to get the writing done.

Greta considered that the sample would be a truer demonstration of the writer's ability as "if it's something that they've chosen and they are motivated to do then I might get three or four sentences and I can get a much truer idea of where they are."

In another interview, Joanna, from School 1, described how popular culture characters influenced children's play and learning. She related how year 1 children talked about these characters at news and sharing time saying "Popular culture affects how they dress, how they do their hair, it affects often how they talk." She made it clear how the children in her class based their play on television programmes or computer games, and the "news" they shared with the class and their dramatic play at interval were in a sense a rehearsal for their writing.

You see games in the playground too. They will be the characters. Sometimes they will play Transformers and you can get them to write about that because they will say it for news, what they did at lunchtime and you can say that would be a really good story to write about today.

Joanna further explained the value of this as "...more relevant because it's a personal recount and it is something they have enjoyed and it's still fresh in their minds. It's usually after lunch that we write in the juniors." In her class the selection of topics for writing and rehearsing of ideas was shared with children's parents. Joanna stated "Some children will think about what they will write about in the morning. We encourage parents to talk about that when they are eating their breakfast so they have something in their mind or they will come in with something they've done or seen just now." In the back of each child's writing book was an individualised goalsheet with "I am learning to..." and "I can..." followed by the writing skills that she had written in for each child. In her class children practised these skills when they were writing about their self-selected popular culture topics.

Joanna also described a multimodal approach to engage and motivate her children in writing when she showed her class the theme song from Ben 10 on Youtube. This visual stimulus led to children discussing then writing about the characters. Joanna commented: "It's fun and they are glued from the moment you turn it on." She explained:

The children could tell me about the characters because all the characters were introduced during the song, not verbally but with the pictures. That jiggled their memory of who everyone was and after that we had a discussion, I have been doing effective teaching strategies, think pair share, write about a character, talk to a partner, what do you know about that person and we report back.

Her explanation of the strategies used in her classroom illustrates a belief in a social approach to learning.

Joanna clearly understood her children's interest in the Ben 10 characters and how to enter the spirit of the genre to motivate their writing. When one girl, Irene, wrote that she liked Gwen, who is Ben 10's cousin, she had not recorded why she liked this character. Joanna explained how she would help Irene to extend her writing:

I said to her I know nothing about this. You've got to pretend that I don't know anything, I'm an alien and I've just come to live in New Zealand and everyone's wearing Ben 10 tee shirts and they need to tell me who he is, describe him.

12.11.10

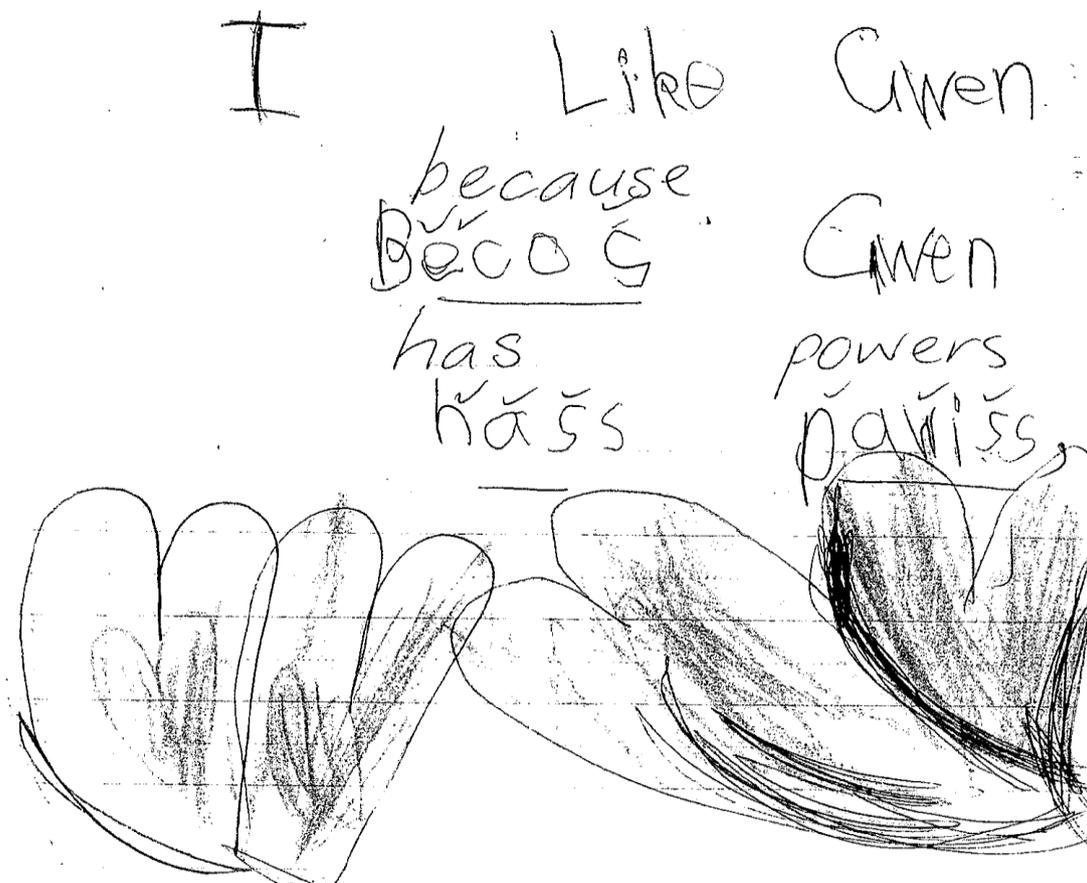


Figure 2: Example of Irene's Writing

Irene returned to her book and added in an explanation of why she liked Gwen "Beco Gwen hass powiss" (because Gwen has powers). She extended her simple three-word sentence to a complex sentence and demonstrated her awareness of sound to letter relationships. Irene illustrated Gwen's powers by drawing two large pink hands to illustrate how Gwen fights her enemies with bursts of pink energy from her hands (see Figure 2).

When teachers were asked in the interviews if they thought that the introduction of national standards in reading and writing would have an influence on the use and acceptance of popular culture in schools, their responses indicated their belief in incorporating children's own knowledge and interest into school literacy. Greta, in School 2, for example, stated that there should be a balance of what is used in schools: "the incorporation of popular culture can motivate and engage children." She explained that any assessment needs to be considered carefully but using popular culture for assessment was just as valid as other topics. The response of Sarah, another teacher in School 2, clearly describes her view of incorporation (McNaughton, 2002). She explained that:

The more links we make to the children's real world the more effective our teaching. For reluctant readers and writers, linking to topics that they find motivating can make a huge difference. Giving children the opportunity to write about things that have been happening to them has been really valuable.

Sarah commented that some of the class writing time was driven by school requirements such as the need to teach explanations, and teachers wanted to give children the opportunity to have their voice.

Although these examples contribute towards informal assessment, some forms of formal assessment are also required. By way of example, Joanna explained that the class programme could include popular culture aspects so when formal assessment is done as part of overall teacher judgment, children have already had experience of the required skills and knowledge.

Discussion

In agreement with Marsh (2006), popular culture is integral to young children's engagement in a wide variety of literacy experiences, some of which are interceded by new technologies, and as a result is an important part of their social practices outside of school. Children are surrounded by characters from popular culture with images, symbols and values, and they are seen on television, in movies, books, and computer games. Ben 10's image, for instance, is seen on children's stationery, tee shirts, duvet covers, reward stickers, as well as party tablecloths, hats and invitations, skateboards and school bags. Children's TV programmes are interspersed with advertisements such as Burger King advertising Ben 10 magnets given free with children's meals. Marsh (2006) argues that popular culture provides a choice of material that young children find engaging and it has the "potential to motivate students who might otherwise think their particular cultural interests are excluded from the curriculum" (p. 160). The survey results indicated that many teachers are taking into consideration children's out-of-school literacy practices and making use of these in their school's literacy curriculum programme.

The descriptions from these teachers of how they encourage children to write about their out-of-school interests in order to "make connections" to school literacy are examples of McNaughton's (2002) incorporation described in the literature review and they are consistent with the socio-cultural approach of the *Effective Literacy Practice* handbooks (Ministry of Education, 2003, 2006). The assessment paper mentioned earlier (Ministry of Education, 2010a) requires that "the approaches and tools will be chosen to collect a rich and reliable picture of student achievement that supports strategies to improve achievement" (p.36). If this is taken into consideration within a socio-cultural perspective of learning then teachers can draw on examples such as these in writing in addition to their selection of formal assessment tools.

All these instances of children's writing described by the teachers in the survey and interviews could provide teachers with useful material to help them form the required professional qualitative judgments across the range of formal and informal assessment through conversations, observations, tasks and tools described in the Ministry of Education's (2010a) position paper on assessment (see Figure 2, p. 8).

During the interviews with the four teachers it emerged that each had a detailed understanding of their young pupils' popular culture. Their interest and knowledge were apparent in detailed descriptions of, for example, the characters, television programmes and computer games that children they taught talked about, played, drew, and wrote about. Their beliefs about teaching and learning seemed consistent with a socio-cultural approach to meet the requirements of the curriculum.

All the teachers in this research are working with young children who are still in the early stages of learning to decode and encode in English. It might be assumed that it is more appropriate at this stage of schooling rather than later to make links to children's out-of-school knowledge in school literacy programmes as children acquire the skills and knowledge to use "reading and writing as interactive tools to enable them to learn in all curriculum areas" (Ministry of Education, 2009, p. 13). However, schools and teachers

need to make learning relevant and interesting for learners right through the years of compulsory education, and incorporating their expertise in popular culture is a way to achieve this.

In this paper we asked the question to what extent do primary teachers in years 1 to 3 incorporate children's interest and expertise in popular culture as they assess literacy. Primary teachers in New Zealand are expected to take the theoretical perspective offered in the Ministry handbooks (Ministry of Education, 2003, 2006) of a social practice interpretation of literacy. However, alongside this go the detailed skills and knowledge requirements for each level of the *Reading and Writing Standards* (Ministry of Education, 2009) and *The Literacy Learning Progressions* (Ministry of Education, 2010b). The implication for practice is that there is a challenge for teachers to reconcile these two demands in ways that make literacy learning meaningful, motivating and successful for children while meeting the requirements of the curriculum. The majority of teachers in our survey reported that they incorporated aspects of popular culture as they taught and assessed literacy in their classroom programmes. Most of the examples linking popular culture and literacy were of children's own choices for writing, and teachers described how children were often involved in dramatic play at interval before writing. Although some teachers did not make links to popular culture in class literacy programmes, the majority believed that it was appropriate to use these topics to contribute to their overall teacher judgments of their children's literacy.

Conclusion

It is possible that the introduction of national standards in New Zealand schools will lead to literacy teaching becoming focused on skills in a way that reflects Street's (1984) autonomous paradigm. The results of this survey and the case studies suggest that literacy learning and teaching in New Zealand schools can maintain a socio-cultural perspective and all the teaching of the required skills and knowledge by making links to children's out-of-school knowledge and interest in popular culture.

There is a need for future research in the middle and senior primary school to investigate how teachers capitalise on children's interest and expertise in popular culture in school literacy programmes as they assess children's literacy and form overall teacher judgments. The question remains to what extent can this contribute to the "professional qualitative judgments" both "formal and informal" that are required of teachers (Ministry of Education, 2010a, p. 14).

References

- Dickie, J. (2008). *An investigation of sites, uses and practices for literacy in the lives of Pasifika students*. (Unpublished doctoral dissertation), Victoria University of Wellington, New Zealand.
- Dickie, J. (2011). Samoan students documenting their out-of-school literacies: An insider view of conflicting values. *Australian Journal of Language and Literacy*, 34(3), 247-259.
- Dyson, A. H. (1997). *Writing superheroes: Contemporary childhood, popular culture, and classroom literacy*. New York: Teachers College Press.
- Dyson, A. H. (2001). Where are the childhoods in childhood literacy? An exploration in outer (school) space. *Journal of Early Childhood Literacy*, 1(9), 9-38. doi:10.1177/14687984010011002
- Dyson, A. H. (2003). Popular literacies and the "all" children: Rethinking literacy development for contemporary childhoods. *Language Arts*, 81(2), 100-109.
- Gee, J. (1996). *Social linguistics and literacies: Ideology in discourses* (2nd ed.). London: The Falmer Press.
- Heath, S. B. (1983). *Ways with words*. Cambridge: Cambridge University Press.

- Lankshear, C., & Knobel, M. (2003). *New literacies: Changing knowledge and classroom learning*. Buckingham, PA: Open University Press.
- Luke, C. (1997). *Technological literacy*. Canberra: Language Australian.
- Makin, L., Jones Diaz, C., & McLachlan, C. (2007). *Literacies in childhood: Changing views, challenging practice* (2nd ed.). Marrickville, NSW: Elsevier Australia.
- Marsh, J. (2005). Ritual, performance and identity construction: Young children's engagement with popular culture and media texts. In J. Marsh (Ed.), *Popular culture, new media and digital literacy in early childhood* (pp. 28-50). London: RoutledgeFalmer.
- Marsh, J. (2006). Popular culture in the literacy curriculum: A Bourdieuan analysis. *Reading Research Quarterly*, 41(2), 160-174. doi:10.1598/RRQ.41.2.1
- Marsh, J. (2009). Writing and popular culture. In R. Beard, J. Riley, D. Myhill, & M. Nystrand (Eds.), *The sage handbook of writing development* (pp. 313-324). London: Sage.
- Marsh, J., & Millard, E. (2000). *Literacy and popular culture: Using children's culture in the classroom*. London: Paul Chapman.
- McNaughton, S. (2002). *Meeting of minds*. Wellington: Learning Media.
- Ministry of Education. (2005). *Education standards act*. Wellington: Learning Media.
- Ministry of Education. (2003). *Effective literacy practice in years 1 to 4*. Wellington: Learning Media.
- Ministry of Education. (2006). *Effective literacy practice in years 5 to 8*. Wellington: Learning Media.
- Ministry of Education. (2007). *The New Zealand curriculum*. Wellington: Learning Media.
- Ministry of Education. (2009). *The New Zealand curriculum reading and writing standards for years 1 - 8*. Wellington: Learning Media.
- Ministry of Education. (2010a). *Ministry of education position paper: Assessment*. Retrieved November 5, 2010, from <http://www.minedu.govt.nz/~media/MinEdu/Files/TheMinistry/AssessmentPositionPaperSep2010.pdf>
- Ministry of Education. (2010b). *The literacy learning progressions*. Wellington: Learning Media.
- Morrell, E. (2002). Toward a critical pedagogy of popular culture: Literacy development among urban youth. *Journal of Adolescent & Adult Literacy*, 46(1), 72-77. doi:10.2307/821822
- Razfar, A., & Gutierrez, K. (2003). Reconceptualizing early childhood literacy: The sociocultural influence. In N. Hall, J. Larson, & J. Marsh (Eds.), *Handbook of early childhood literacy* (pp. 34-47). London: Sage.
- Robbins, J. (2005). 'Brown paper packages'? A sociocultural perspective on young children's ideas in science. *Research in Science Education*, 35, 151-172. doi:10.1007/s11165-005-0092-x
- Street, B. (1984). *Literacy in theory and practice*. Cambridge: Cambridge University Press.
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Yin, R. (2009). *Case study research: Design and methods* (4th ed.). Los Angeles, CA: Sage.

Promoting Information Literacy Development by Assessing the Research Process in First-year Tertiary Assessments

Angela Feekery & Lisa Emerson
Massey University, Palmerston North Canterbury

Email: A.J.Feekery@massey.ac.nz; L.Emerson@massey.ac.nz

Abstract

One of the challenges students face as they enter tertiary education is having to find, filter and engage with information sources needed to complete assessments. Head and Eisenberg (2010) showed that, while today's students seem to have developed good information seeking behaviours, they struggle with the open-endedness and overabundance of electronic information available in the digital age. Source evaluation and usage remain problematic – and most forms of summative assessment do not allow for student reflection on the learning process or source usage. This paper addresses the challenges of source evaluation and use in written assessments. It examines innovative forms of assessment which enable students to think critically about source usage and evaluation, and which provide academic staff with insight into the information literacy skills and challenges of their students. It shows that, while information search skills may be most effectively taught by library staff, deeper aspects of information literacy (such as source evaluation and justification of usage) are discipline specific and therefore most effectively embedded in the disciplines through carefully crafted forms of formative and summative assessment.

Introduction

Since the 1990s, the amount of information available in both traditional forms and electronically has increased exponentially. In fact, it is estimated that “the sum of human knowledge will double every 73 days by 2020” (Breivik, 2007, in Andretta, Pope & Walton, 2008, p. 46). Due to the overabundance and open-endedness of electronic information, Head and Eisenberg (2010) concluded that “research is one of the most difficult challenges facing students in the digital age” (p. 2). Thus, one of the major challenges for students entering tertiary education is having to find, filter, evaluate and engage with information sources needed to complete assessments: that is, to become information literate.

Definitions of information literacy are largely consistent and are an important consideration for tertiary education. The American Library Association Presidential Committee's (1989) widely cited definition of information literacy states “information literate people know how to find, evaluate and use information effectively (para. 21) to solve a problem or complete a task. Information literate students can critically engage with content, conduct extended research independently and collaboratively, and become reflective, self-directed learners (Bundy, 2004; Pope & Walton, 2006). The four key stages of information literacy are: identifying an information need; finding appropriate information; evaluating information; and using and synthesising information. In an academic context, the latter two stages are more closely linked to the discipline than the library, as the discipline determines what information and knowledge is valued, and the instructor determines and controls the way the information will be used when setting assessments. To be successful learners who can critically evaluate information, students must develop an understanding of how knowledge is created within their discipline (Grafstein, 2002). Therefore, developing students' information literacy skills is a critical concern for tertiary instructors.

This paper, part of a longitudinal study on information literacy development in the disciplines, discusses the importance of providing opportunities for students to become aware of, and reflect on, their learning process to aid the development of information

literacy and academic writing skills. In particular, the paper focuses on two key issues in relation to information literacy – how to integrate deliberative exploration of the research process into formative and summative assessment, and how to model effective source selection and evaluation – and details two formative assessment strategies which are designed to support students' acquisition of information literacy skills.

Our research shows that instructors of first year courses (and beyond) can explicitly help students' successful transition into tertiary study by creating carefully crafted forms of formative and summative assessment that focus on the research process and information literacy skills development appropriate for each content domain.

Brief Overview of the Research

The formative assessments outlined in this paper have been used as interventions in a larger participatory action research project aimed at identifying how the development of information literacy academic writing skills can be embedded¹ across a full undergraduate programme within the context of a specific discipline at Massey University. Lecturers in the participating programme (hereafter referred to as 'participating lecturers') volunteered to be a part of the research due to concerns about students' source selection strategies and academic writing skills across the four-year programme. The two assessments outlined here have been implemented in two core first year papers of the programme. The assessments were aimed at encouraging students to think about the academic research and writing process, with a particular focus on source selection, evaluation and use.

A key focus of our research is to identify and implement ways students can appropriately synthesise information from a variety of primary and secondary sources, and relate the information to assessment tasks. Integrating new knowledge successfully into writing is an important component of academic information literacy development. The interventions into the participating programme include a combination of changes to assessment, and the inclusion of classroom activities and resources that promote the development of information literacy and writing skills.

Data have been collected across two semesters (action research cycles) from both staff and students via meeting notes, focus groups, journals, staff interviews, and anonymous surveys. The data were then manually coded and analysed for emerging themes, and are being used to evaluate the impact of the interventions. In this paper, we will draw on this data set to support our discussion of student problems with information literacy, as well as to provide staff and student responses to the information literacy strategy trialled through our research.

Impact of Assessment and Feedback on Learning

Our research shows that common approaches to assessment lead to two problems in relation to information literacy: faculty underestimating student engagement with information literacy skills, and students undervaluing the importance of developing and demonstrating these skills.

Assessment is an essential part of quality learning and teaching, which strongly influences how students approach the learning process. Carless, Joughin, and Liu (2006) point out that "assessment impacts on what content students focus on, their approaches to learning and their patterns of study" (p. 2) and that assessment changes students' behaviour more than anything else we may teach. Student behaviours that often cause concerns for tertiary instructors include: a perceived lack of time spent researching and reading widely on the topic; seemingly random source selection; the over-reliance on the internet as a

¹ A programme where information literacy skills development is *Embedded*, means the curriculum design is one where students have ongoing interaction and reflection with information, rather than extra-curricular classes and self-paced packages (*Generic* and *Parallel* – complement the curriculum) or separate classes and packages that are part of the curriculum (*Integrated*) (Bundy, 2004, p. 6).

source of information for scholarly tasks; and poor writing ability, including frequent copy and paste, and plagiarism (Brabazon, 2007; Emerson, Muirhead, & Stevens, 2008; Planning Staff, personal communication, 2010; Walden & Peacock, 2008).

These concerns centre on a lack of information literacy skills, yet these skills are rarely directly addressed through the assessment process. Instructors often have limited understanding of how students found sources, or they may see the list of URLs in the reference list as evidence that students failed to look beyond Google. However, reference lists do not indicate how many other sources were rejected during the research process. Our research showed that faculty underestimated the extent of students' engagement with source selection; in terms of time spent researching versus writing, a typical student comment was "I spend too much time looking for information in relation to actually reading and writing it" (Student focus group interviews, personal communication, June 9, 2011).

Assessment can be used to both support learning and measure student performance in relation to information literacy. Formative 'assessment for learning' can be designed to help students learn by both identifying errors and reinforcing correct understanding (Bhattacharya & Jorgensen, 2008). Yet, much of the teaching at university tends to concentrate on content instruction assessed via summative assessment of learning, which sums up the students' achievement, but may not actively promote learning (Sadler, 1984, cited in Pickford & Brown, 2006). Instructors can overcome the challenges of source selection, evaluation and usage by changing their approach to assessment and assessing both the process and the product, to make assessment for learning a larger part of their teaching.

Since students are highly motivated by grades (Dolan & Martorella, 2003; Proctor, 2006) and summative assessment generally does not support or accurately reveal students' engagement or competence in relation to information literacy, students tend to undervalue the development of these skills. Formative assessments that directly support and focus on information literacy are therefore crucial: by making skills development part of the assessment, the tasks are "raised to a level above merely a means to an end" (Dolan & Martorella, 2003, p. 1324). Formative assessment strategies to support information literacy development need to include learning outcomes focused on developing information literacy skills alongside subject-specific knowledge and proficiency (Carless et al., 2006). Student feedback on the formative assessment approaches in our research support this observation: "The [formative] interventions brought the writing of this paper to a higher level and made me more motivated to do them. It seemed like there were expectations on us and a true place for improvement for us as learners" (Student survey, 13 October, 2010).

Deliberative Exploration of the Research and Learning Process in Formative and Summative Assessment

While students have much freer access to information via the multiple search engines available to them (Head & Eisenberg, 2010), the intellectual access (Barton, n.d.) to information that contributes to deeper learning and knowledge development in the disciplines is often underdeveloped in first year students.

Perceived Problems Relating to Information Literacy Acquisition

Our research revealed faculty perceptions concerning students' writing and information literacy skills that supported observations in the literature (Andretta, 2006; Brabazon, 2007; Lea & Street, 1998) i.e., that today's students: "want to be spoon-fed and out of here as fast as possible"; "lack highly developed analytical skills" needed to become independent learners; and have "only passing levels of written communication skills" (Staff interviews, personal communication, 2010).

These concerns expressed by the participating lecturers are not new concerns, (Reynolds, Ker, & Day, 1996), nor are they perceived to be limited to only a few students (Wall, 2006). The gap between high school preparation and university demands appears to be widening, and more students seem unprepared for university study (Angier & Palmer, 2006; Callahan & Chumney, 2009). Brabazon (2007) has observed that a larger proportion of students “require greater guidance, attention and commitment from teaching staff to pass a course” (p. 15), yet still produce inadequate work. Participating lecturers have also identified a “long tail of students that seems to be getting bigger each year who have just very poor writing skills” (Staff interview, personal communication, October 21, 2010). While some students “display a profound curiosity in their topics”, a large proportion treat their courses as “a very utilitarian exercise ... a way of getting a ticket through to a professional career” (Staff interview, personal communication, February 3, 2011). Some lecturers also argued that the step-by-step approach to learning fostered by the NCEA negatively impacts on students’ approaches to learning at university.²

In addition, some faculty perceive problems with writing as remedial or deficit (Crozier, 2007; Murray & Kirton, 2006; Radloff, 2006), as opposed to skills that haven’t been taught yet. Comments from our staff interviews clearly showed the perception that students with issues in research and writing should be referred to, or independently seek support from the university’s student support services and library. However, this relies on weaker students actually being aware of their problems. While these services can and do help students focus on finding sources and writing development, librarians and the writing support advisors are often unfamiliar with the content requirements of the courses (Radloff, 2006), and lecturers’ expectations of the assessments. Some students and lecturers in our research voiced concerns over the usefulness of the advice given by student learning advisors.³

Further, while information literacy is an essential component of the research process (Andretta, 2006), and even though academics believe information literacy and the ability to communicate effectively are essential skills for students, the development of these skills “is an implicit assumption but not an explicit learning outcome” (Staff interviews, personal communication, October 29, 2010). Head & Eisenberg (2011) argue “many – not all – educators are failing to teach students how to navigate a vast wilderness of information – to discern what they can trust, edit out what is unnecessary, redundant or unreliable, and focus on what they really need” (para. 3), and McGuinness (2006) suggests that academics do not see skills development as part of their teaching responsibilities. However, participating lecturers stated their responsibility for teaching academic skills is either something they have not considered, or which they feel unqualified to do themselves.⁴ One staff member interviewed suggested that the research-based tasks being set require students to develop information literacy and good communication skills, and that feedback, if used effectively, will help students improve over time.⁵ This comment

² “in NCEA, as long as you know the steps to go through you’ll get your achieve status...[Students] will say ‘what do I need to do to pass this?’. That’s an entirely wrong way to look at it from a learning perspective, but at the end of the day their goal is to get the degree, not necessarily to learn” (Staff interviews, personal communication, October 22, 2010).

³ One student commented “they made me change my whole introduction... and then I got my feedback saying that I lost marks because I didn’t state how I had it originally (Student focus groups, personal communication, October 14, 2010). Another felt that “they just work on what they think you should be working on rather than what I want” (Student focus groups, personal communication, October 13, 2010). One lecturer also commented that “the student learning centre does depend very much on generalised models. It depends on working out sort of pre-thought-out solutions when in fact each of those students actually has a slightly different problem” (Staff interview, personal communication, October 22, 2010).

⁴ “teaching skills is not something that I have thought consciously about before” (Staff interviews, personal communication, October 19, 2010); “students need to get support and advice from people who are qualified to teach writing (Staff interviews, personal communication, October 12, 2010).

⁵ “through a sequential process over time the students are getting feedback, and those that want to get the feedback and ... take it seriously will improve over the four years. Given the amount of assessment they are doing, they should be improving and in general they do” (Staff interviews, personal communication, February 12, 2011).

supports McGuinness' (2006) observation that faculty believe students will learn information literacy and academic writing skills by being exposed to the skills they need to complete assessment. While 'learning by doing' is a valid way of developing skills, it is not useful if students are engaging in poor and erroneous practice without sufficient instructor guidance and consistent, detailed feedback along the way.

These three sets of perceptions – that students lack willingness to learn, that failures in academic writing and information literacy are issues of remediation, and that academic staff should only have a limited role in developing students' information literacy and writing skills – need to be challenged if the current situation, of students struggling, with minimal aid, to learn essential skills, is to change.

The Solutions

A strong focus on developing information literacy and academic writing skills embedded within the disciplinary context is an effective strategy to enable students to develop essential academic skills. It is important for tertiary instructors to re-conceptualise writing problems, not as remedial, but rather as developmental (Angier & Palmer, 2006); that is, tertiary academic conventions need to be taught at tertiary institutions via integration in either general or discipline-specific programmes (Husain & Waterfield, 2006). While some students may be aware of the academic research and writing process, many others will enter their programmes of study without any formal introduction to academic research and writing, and will struggle to learn these skills on their own. Emerson (2008) found that first year university students had little understanding of the difference between quoting or paraphrasing, or how to acknowledge sources effectively. This failure to acknowledge and incorporate relevant material appropriately into their writing can lead to unintentional plagiarism (Ahmed & McMahan, 2006). Therefore, instructors must not only teach the content of their subject area, but also the skills needed to comprehend and communicate this content effectively (Angier & Palmer, 2006).

First-year students need to be made aware of the academic research and writing process before they can complete assessments successfully. They often lack any disciplinary expertise (University of Washington Libraries, 2010), and should be considered novices on a journey to becoming scholars (Chanock, 2001; Dreyfus & Dreyfus, 1980; Husain & Waterfield, 2006), which means they need to be exposed to "repeated opportunities for seeking, evaluating, managing and applying information gathered from multiple sources and obtained from discipline specific research methods" (Bundy, 2004, p. 6). This includes having opportunities to think about both what they are learning, and how they are learning it. Content instructors play an important role in coaching students through the research process through to writing papers, and they need to focus substantially on helping students to learn and practise research skills (Head & Eisenberg, 2009b, 2010).

The selection of quality of information for assessments also needs to be emphasised from the outset of tertiary studies. Because the first year of university study provides a knowledge base to be built on throughout the degree, instructors need to ensure that the information that is providing a foundation for that knowledge development is coming from reliable sources that are valued in their disciplines. Instructors also need to accept that internet search tools, especially Google, are part of the current learning reality, and need to be utilised effectively rather than dismissed outright (Godwin, 2006; Hepworth, 2006). However, "to access and download internet content is not to learn" (Wall, 2006, p. xiii). If instructors want assessments where the argument is supported by strong, scholarly evidence, then they have to encourage and explicitly teach students how to go beyond the first few web pages on Google. When the internet is used, students should provide a reflective justification of the selection to ensure they have considered the relevance of the source to the task and how it connects to other sources used.

Encouraging a focus on the learning process by having submissions of key tasks (i.e., formative assessments) during task completion can help students identify the stages in

the process that are often inadequately addressed. Head and Eisenberg's (2009a) research confirmed what many instructors believe: the majority of students did not start an assignment until two to three days before it was due (p. 7). Students in our focus groups alluded to this as well; those who started the assessments more than a week in advance were the exception. Our first year students seemed to appreciate earlier formative submissions; one student commented "more deadlines for doing many tasks makes me work" (Student survey, personal communication, October 13, 2011). However, instructors need to be careful that they are not 'spoon-feeding' old habits; therefore, the level of scaffolded support for familiar assessment tasks should be reduced at the higher levels of the degree.

To ensure that students are able to focus on the learning process, a balance of formative learning tasks throughout coursework and assessment, particularly in the first year of tertiary study, is essential. Pickford and Brown (2006) suggest that "a shift from focusing on what is taught to focusing on what is learned makes it inevitable that assessment of process is emphasized" (p. 99). This should enable students to learn the skills they need to become successful learners who employ good processes and have the knowledge and ability to produce outcomes relevant to the discipline. Students are also less able to plagiarise if they are asked to demonstrate work in progress, and are assessed for their efforts during the research process (Bhattacharya & Jorgensen, 2008; Emerson, 2008; Pickford & Brown, 2006). Furthermore, if instructors can see how students come to their conclusions, and what sources influence the opinion they hold, they can have a better idea of the gaps in students' learning and their teaching.

However, because formative assessment can increase the workload of both instructors and students,⁶ "instructors need to carefully consider when to assess the process and what benefits it has to the students to focus on assessing process rather than the outcome achieved" (Pickford & Brown, 2006, p. 100). Carless et al. (2006) state instructors must face the challenge of creating assessment that improves the quality of student learning without increasing the workload of both staff and students unnecessarily. When formative assessment activities are based on key tasks in the research process, requiring students to record and be reflective about their own learning should help them discover ways to work smarter, not harder. Also, the marking workload on the final assessments should be reduced as instructors are being presented with better quality assessments. However, we do need to ensure that assessments contribute to a specific learning outcome. The learning outcome that was added to one of the courses participating in our research was 'to enable you to have the ability to find, evaluate and use information effectively to enhance your own learning and personal development'.

Assessment that focuses on the research process should be evaluated in terms of the level of reflection students show on how content and skills are being learned during the assessment process. Pickford and Brown (2006) believe that "encouraging students to reflect on experiences and to learn from them is unquestionably a useful formative exercise" (p. 107) when used in conjunction with other forms of assessment. Students can reflect on how much effort they put into a task, and identify problems within their process that are hindering successful learning outcomes. If they also get feedback on their process during the process itself, either from instructors or peers, then the process becomes more relevant in the overall assessment process.

The Example: I-map (process) Assessment

The challenge in assessing the research process is developing an assessment tool that provides an accurate overview of the process, but that can also become a useful tool for future assessment tasks. The i-map (see Figure 1) fulfils both these functions.

⁶ "we have to be really careful about not overloading the students, and to get them to do work, you have to put assessment on it" (Staff interview, personal communication, October 19, 2010).

The i-map is a visual depiction of the entire research process (Emerson et al., 2008; Walden & Peacock, 2008), from understanding or forming the question to justifying source selection, and reflecting on the process. Head and Eisenberg's (2010) findings suggest that students need to be held accountable for the research they conduct and that as part of any research-based assessment, they should "substantiate their research strategy, evaluation and selection of sources and show hard evidence of critical thinking about information" (p. 39). The i-map incorporates this strategy, and provides an "assessable outcome that rewards process, including information literacy skills, rather than product" alone (Walden & Peacock, 2008, p. 133).

With its emphasis on the research process, the i-map has become an important assessment tool for two first-year core writing papers at Massey University over the past few years (Emerson et al., 2008). In both papers, a series of scaffolded in-class activities including brainstorming, developing questions to investigate, identifying keywords for information searches, creating thesis statements, and evaluating and justifying sources, contribute to the development of the i-map during the research, planning and writing process. Feedback on these activities is provided through the teaching and peer assessment conducted within the courses. While the i-map itself then becomes a summative account of the research process that can be assessed, the activities used to create the i-map are effectively contributing to development of information literacy and writing skills.

Vivisection is an inevitable aspect of modern life

Focus Questions:

Is vivisection an inevitable aspect of modern life? Is vivisection ethically acceptable? How should the minister approach the subject of vivisection when confronted by the media?

Individual Roles decided:

Tom - Leader & co-editor
Bridgette - Co-editor
Samantha - I-map
Byron - I-map
Lilian - Secretary and team analysis

-Team goals set
-Aims for next meeting made
-General research begins from each group member



Discussions on interview to be done:

Tom - anti-vivisectionist
Sam - director of NZAVS
Byron - member of ethics committee

Paragraph decisions and allocation:

Value of life - Tom
Efficiency - Sam
Legality - Bridgette
Freedom - Byron
Education - Lilian

Interview questions made:

Who are you and what is your involvement with vivisection?
Do you think that people in NZ are aware of vivisection and what it is?
What are some alternatives to vivisection?
What is your stance on the subject of vivisection?
Do you agree with the rules and regulations surrounding vivisection?
Provided you are you aware of them?
Do you think there is enough money invested into animal experimentation?
Are New Zealanders educated about vivisection?
How do you educate people about vivisection?
Is vivisection done to all different types of animals (higher order/lower order)?
Do you think the benefits are worth the consequences?
What process do you need to go through to get consent for vivisection?
Do your experiments provide good/useful results?

Keywords used during searches:

Vivisection, ethics, rights, utilitarian values, respect, efficiency, education, regulations, laws, rules, culture, religion, anti-vivisection, statistics, New Zealand



Search strategies of group:

- Library search:** The use of the Massey University library resulted in many relevant books. Of these, two which were particularly useful for Byron's, Bridgette's, and Lilian's paragraphs includes:
 - Bourke, D., & Edan, M. (2003). *Lifting the Veil of Secrecy on Live Animal Experiments*. Auckland, Wellington, Christchurch: SAFE.
 - Balcombe, J. (2000). The use of animals in higher education: Problems, alternatives and recommendations. Washington: Humane society press.
 This two books had great relevance for information about education and the legalities of vivisection. The use of books is also a fairly credible and reliable source to use.
- Journal Articles:** To find useful and credible articles, Academic Search Elite, which is part of Massey University's Journal Catalogues, was used. The articles produced had good information and are peer-reviewed which boosts their credibility. Some articles found which were found to be relevant for Sam's and Tom's include:
 - Cuthill, I. C. (2007). Ethical regulation and animal science: why animal behaviour is not so special. *Animal Behaviour*, 74, Issue 1, 15-22.
 - Tatchell, P. (2004). Why animal research is bad science. *New Statesman*, 133, Issue 4700, 18-19.
 - Walport, M. (2006). Cure versus cruelty. *New Scientist*, 190, Issue 2555, 6.
 - Festing, S. (2005). The animal research debate. *The Political Quarterly Publishing Co*, 76, Issue 4, 688-672.
- Web Search:** The use of the internet and Google searches was more minimally used than the previous two strategies, however, it was still useful in bringing our attention to some relevant websites. Included in this was the New Zealand anti-vivisection society website (www.nzavs.org.nz). This site had relevant information and was a useful in background information on NZAVS before interview of director.

Interviews:

- Three interviews were done, each of a person with an involvement with vivisection. It was made sure that each of the three stances on vivisection were covered.
1. A tertiary institution vivisectionist
 2. Phil Clayton, director of New Zealand Anti-vivisection Society
 3. Kevin Stafford, member of the Massey University animal ethics committee
- From doing people with very differing opinions on vivisection this allowed for a variety of perspectives to be examined.
- Some useful information which came out of these includes:
- Some of the vivisection procedures which are done in New Zealand
 - The three R's relating to vivisection: reduction, replacement and refinement
 - Both benefits and consequences of the research: Vivisectionist: easier and safer than testing on humans. Anti-vivisectionist: Animals not good models for humans and products are harming ourselves and our families.
 - Process to be undertaken to be able to do vivisection in New Zealand.
 - Where the funds for vivisection come from in New Zealand
 - That the New Zealand public has the wrong idea about vivisection
 - Scientists prefer to use other methods of research because animals cost too much both in price and ethics

Breakdown of references used in report by group:

- 1) Paixão, R. L., Schramm, F. R. (1999). Ethics and animal experimentation: what is debated? *Cad. Saúde Pública*, 15 (1), 98-110.
-from a Brazilian journal which has been translated
-explains different sides of the ethical debate on vivisection
-strong history emphasis, how it came about
-used article to find information on different philosophies of vivisection
-also used as useful information for the three R's (reduction, replacement and refinement), which are points that came up during a completed interview.
- 2) Orlans, F. B. (1993). *In the name of science: Issues in responsible animal experimentation*. New York: Oxford University press.
-useful for information on the following philosophies: animal treatment, how animals could be used as long as there is a purpose and animals would not be harmed, and also that animals should not be used for science because of their have basic moral rights to be treated with respect
-large section about animal rights was helpful also
- 3) Resnik, D. B. (1998). *The ethics of science*. London and New York: Routledge.
-from study guides and useful as defines the term utilitarian which is the main angle of our report.
-useful also for information on the standards of ethical conduct in science which led to the paragraphs for discussion being chosen. Included in this is efficiency, legality, freedom, education and respect for subjects
- 4) Cuthill, I. C. (2007). Ethical regulation and animal science: why animal behaviour is not so special. *Animal Behaviour*, 74, Issue 1, 15-22.
-from the Massey article database, peer-reviewed articles
-looks at all issues surrounding vivisection including consequences to animals, utilitarian view, and the reasons it is used for medical research
-used as its information on efficiency, in particular, utilitarian perspective was useful
- 5) Tatchell, P. (2004). Why animal research is bad science. *New Statesman*, 133, Issue 4700, 18-19.
-from the Massey article database, peer-reviewed articles
-very useful information surrounding the anti-vivisectionists opinions as it explains the reasons why animals can not be good models for humans in medical terms
-used article to show this viewpoint in efficiency paragraph
- 6) Walport, M. (2006). Cure versus cruelty. *New Scientist*, 190, Issue 2555, 6.
-from the Massey article database, peer-reviewed articles
-this article came from the opposite viewpoint in that it supported the use of vivisection
-it was used to give benefits of vivisection to back up the medical positives of this research
- 7) Regan, T. (2005). *Empty Cages: Animal Rights and Vivisection*. Oxford University: Black Well Publishing.
-from the Massey article database, peer-reviewed articles
-useful information from this text includes: the benefits of vivisection, who supplies the funding in New Zealand, the ethics involved, and how the freedom of scientists reflect the research which is carried out.
- 8) Bourke, D., & Edan, M. (2003). *Lifting the Veil of Secrecy on Live Animal Experiments*. Auckland; Wellington, Christchurch: SAFE.
-from the Massey University library
-this book has a very bias view of vivisection which shows the negatives, major concerns, and where it occurs, with strong examples of badly done vivisection
-it outlines the rules and regulations surrounding vivisection and it has in depth information about institutions and organisations that participate in this research
- 9) Balcombe, J. (2000). *The use of animals in higher education: Problems, alternatives and recommendations*. Washington: Humane society press.
-from the Massey University library
-this book gives a view of harming animals for education as undesirable and it also looks in depth at the emotional tolls on students of them dissecting live and deceased animals.
-alternatives to live animal dissection are offered by the author and he believes students should be fully involved in decision making and be aware of ethics.

Timelines for completion:

- References to be emailed to Sam by each member for I-map by 2nd October
- Team role opinions to be sent to Lilian by each member by 8th October
- All written work to completed and sent to Tom for editing by the 11th October
- I-map to be near completion
- 13th October: Peer Review (need draft)
- 16th October = Tutor Clinic. 2.30pm (need all report)

-Addition of media paragraph (Bridgette)

- Cultural angle added (Research—Sam, Add-in—Tom)
- Allocation of final appendices:
 - Team roles—Lilian
 - Critical summaries—Sam, Bridgette, and Byron
- Bringing together of all paragraphs for editing
- Conclusions and recommendations to be written



Structure of essay and opening sentences:

1. Introduction:

2. Discussion:

- 2.1 - Value of life
 - The value of the lives of animals is arguably the most contentious debate on the subject of vivisection. The question of whether humans are superior to animals, and whether animals can feel emotions and pain are central to this debate.
- 2.2 - Efficiency
 - Another ethical principle which is highly important in this debate is the principle of efficiency. Efficiency involves looking at how scientists use their resources and their subjects in experimentation, and whether it is efficient or not. Given that scientists have limited economic, subject, and technological resources it is important that these resources are not wasted whilst achieving their aims.
- 2.3 - Legality
 - Vivisection, because it uses live animals in experiments, is an area of scientific research that is heavily regulated. There are many regulations regarding what can and cannot be done as scientific research. These regulations are based on reducing animal abuse and misuse, and may be based on either scientific, moral or ethical principles.
- 2.4 - Freedom
 - Due to the laws and regulations surrounding vivisection, freedom for scientists can be very restricted. Ethically, scientists should be free to conduct research on any problem or hypothesis, however this principle of freedom needs to be balanced with other concerns.
- 2.5 - Education
 - An important part of research is to share newly discovered information in order to educate other scientists. This ensures that the information from the research is spread throughout the scientific community and contributes to existing knowledge. Scientists should also educate and inform the general public about scientific breakthroughs and about ongoing research.

3. Conclusions:

Each links back to its relative discussion point its relevant recommendation which follows.

4. Recommendations:

Each of the recommendations relates back to a specific conclusion which has been discussed earlier in report.

Figure 1: Example of an I-map

The i-map was a new concept for the lecturers in the participating programme. One first year lecturer involved in our research trialled the i-map as part of the essay assessment, and provided instructions of what to do. Students requested i-map models as this type of assessment tool was new to them as well.

Unlike the use of the i-map in the writing courses, where activities are used to support the i-map development, in the participating course, the i-map came in at the end with the essay. While instructions and models were provided, students were largely left on their own to reflect on their process and construct the i-map without feedback along the way. The lecturer recognised that “it was learning for me as well, because I have never done an i-map myself, and had never even seen one before, let alone taught anything about it. So now that I have been through one round, I will be clearer about what it is and what I am expecting to see” (Staff interview, personal communication, December 14, 2010). The lecturer enjoyed seeing the visual depiction of the students’ journey as they completed the essay, and decided to use the i-map again in the following semester.

Student responses to the i-map varied but 13 out of 19 students gave it a 4 out of 6 or higher (with 6 being ‘extremely useful’) for usefulness as a tool for helping them successfully complete their essay. The positive comments were that “it helped me plan my thoughts and the relationships between them”, and “identify steps at an early stage” (Student survey, 2011). Only 6 out of 19 students surveyed saw little value in the i-map. The few negative comments related to the students having their own way of researching or that they don’t like visual representations.⁷ The comments against the usefulness of the i-map suggest students may have felt restricted by the models provided and missed Walden and Peacock’s (2006, p. 208) point that a key feature of the process-led forms of assessment like the i-map is that there is no set definition of what they contain. I-maps are designed to acknowledge that people have very different learning styles and may represent them as they think best. Of some concern were the students who completed the i-map retrospectively.

Ways to ensure students are working on the i-map during the research process that will be trialled in the next semester by the participating lecturer include checking the students’ progress on the i-maps during the process, and having an earlier submission to prevent retrospective completion. If time permits, there may also be an opportunity for students to bring their i-maps into the classroom, so they can discuss their process and challenges and help each other identify solutions. These earlier activities using the i-map-in-progress may give the i-map more value if students can see how it is helping their learning.

Addressing the Challenges of Source Selection and Evaluation

Effective source selection and evaluation are important information literacy skills that support the research process. In terms of assessment tasks, source selection is determined by understanding the relevance of a source for a particular task. Appropriate source selection relies on knowing what is relevant, and what to reject (Head & Eisenberg, 2010). However, internet searches for information often produce pages of results where the quality information is hidden among poor quality information. As one student stated, “you try to look for one or two things but you come back with everything but the kitchen sink” (Student focus group, personal communication, October 15, 2010).

The Perceived Problems Relating to Source Selection and Evaluation

Evaluating sources and synthesising them into written assessments is one of the biggest challenges for students. This can be seen in first-year students’ reference lists with heavy reliance on websites as a major reference source (Middle States Commission on Higher Education, 2003). Rather than reading widely to identify themes within the literature to synthesise and develop arguments from, weaker students produce ‘one paragraph, one

⁷ “I would have planned my essay regardless ; “the way I do my brainstorming wasn’t enough – I found I waffled and made up steps” ; “I have my own research process that work for me. They aren’t strictly linear” (Student survey, Semester 2, 2011)

source' essays.⁸ Some students may also rely on quotations or paraphrased ideas,⁹ while lacking a true understanding of the ideas in the text, and believing that as long as they provide a citation, "they have acquired at least the appearance of academic status" (Husain & Waterfield, 2006, p. 27).

Random internet searching for sources and selection based on accessibility are key concerns for tertiary instructors, with one participating lecturer describing student reference lists as a "random shopping list they've managed to accumulate that doesn't tell a story" (Staff interview, personal communication, November 5, 2011). This is impacted by the challenge of using databases (Leckie, 1996), which puts students off seeing these search tools as useful for obtaining information.¹⁰ To overcome this problem, some lecturers have begun to specify the types of sources students should access,¹¹ with the aim of helping students identify relevant scholarly sources in their discipline (Davis, 2003, cited in Middle States Commission for Higher Education, 2003). However, this alone does not help students understand the value of these sources,¹² and further justification of source selection is needed (Brabazon, 2007).

One apparent solution to the issue of random source selection may cause another problem. Providing core readings to students will ensure they are being exposed to relevant information. However, if instructors provide all the readings, they are removing the discovery part of the process, where students explore the range of information available to them. Providing all the reading material may also create issues with controlling the development of knowledge, adding potential bias, and students only learning what the lecturer values (Wade & Moje, 2001, para 9). This means students will be unlikely to extend their understanding of a topic beyond what the instructor presents, and may be challenged in later courses where new instructors hold differing opinions to the previous instructor.

A final concern is students' resistance to reading extended texts. The 2011 Citation Project data showed that 77% of citations come from the first three pages of the source used regardless of the source length (46% from the first page), and suggests that students are not meaningfully engaging with the texts (Jamieson & Howard, 2011). One participating lecturer stated that "students are reading at a very superficial level and then just believing everything that they read and then regurgitating that" (Staff interviews, personal communication, October 19, 2010). The perceived difficulty of reading academic texts is seen as a barrier to first year and second year students using these texts to inform their assessments.¹³ If students are not taught how to read the academic texts provided for them, they won't read them and instead rely on the instructors to identify the key points of the texts in the lectures. This creates a cycle of lack of preparation, which leads to further instructor clarification, which then results in even less preparation (Immerwahr, 2011).

⁸ "a below average student has one source [per paragraph] or maybe one page sometimes just one source" (Staff interviews, personal communication, November 22, 2010).

⁹ One lecturer said students just try "to quote people, and you get that a lot, just a random lot of quotes thrown into sentences just to try to make points, and referencing it, they are not doing it really for any other purpose than to show they have read the article" (Staff interview, personal communication, October 22, 2010). Another stated "they found a phrase and they don't really integrate it into the text and it just sticks out like a door hook" (Staff interview, personal communication, November 5, 2010).

¹⁰ One second year student said "I used to find kind of 5 things that were semi-related to the topic, but then through this year I have started to use more books, but I still find article searching real tricky on the Massey [library] website ... it says you can access it and you go through what it says and you enter stuff in and you just end up with nothing... you end up giving up and just using Google Scholar. I don't really use articles just because I can't function [databases]" (Student focus group, personal communication, October 14, 2010).

¹¹ "I make it compulsory that I really look for 5 scholarly readings in terms of a book or a journal that is not present in the book of readings" (Student focus group, personal communication, November 22, 2010).

¹² One student commented "it is a journal article, so it must be OK" (Focus group, personal communication, September 30, 2010).

¹³ "[first year students] haven't read many journal articles and are in no position to make meaningful judgement about them. They have a hard time just doing the reading" (Staff interviews, personal communication, October 22, 2010).

The Solutions

Being able to evaluate sources effectively comes from understanding what is important in the discipline. Giving students the opportunity to get feedback on sources selected before using them to complete assessments helps students learn how to be selective, and how to determine the true value of their search results. Selecting good, relevant sources, and reading critically (rather than at a surface level) helps students develop a better understanding of the disciplinary knowledge construction and values. Students who are information literate will become better producers of information themselves, because if they are reading relevant and appropriate texts, they will be being exposed to the way to write in the discipline and become familiar with disciplinary discourse early in their academic career.

As mentioned earlier, the provision of readings to students may reduce the use of irrelevant sources in assessments. Our research has revealed that the way readings are provided varies depending on the course, year level, and personal preference of the instructor. The first year courses in our research provided a small set of readings and recommended readings, but most of the searching for information for assessments was done in addition to (or instead of) the material provided for the course. However, for one second and one third year course, the bulk of the readings were provided. When the two lecturers were asked why they provide all the readings for their courses, one answered that he doesn't "want students reading rubbish" and is concerned students won't access the readings themselves. The other lecturer needs a specific group of readings for the content of her course (Planning staff interview, personal communication, 2010).

Since the provision of readings could create problems in terms of reducing the discovery process and controlling knowledge development, instructors need to carefully consider the balance of provided and self-selected source material that is appropriate for their courses. Core readings, recommended readings, and self-selected readings all inform any assessment; the key to successful learning is, firstly, being taught how to read academic texts, and then learning to understand the connections between all the sources used (and what the student already knows), and their relevance to the assessed task (Wade & Moje, 2001). Once students find sources, they need to connect ideas in these sources with what the lecturer has provided for them. But this step is often missed. If students are taught to extend searches using reference lists of relevant, good quality sources (Gaipa, 2004; Lunsford, 2010), they will be able to see how academic knowledge is created by adding to the existing research and knowledge-base of the discipline (Grafstein, 2002). Students will get a much more comprehensive understanding of the issue if they can make connections and start to recognise authors who are being cited in more than one source. Therefore, a useful part of any assessment could be a reflection on the sources used and how they connect to each other.

If instructors want students to learn to value certain information and ideas, then time needs to be spent discussing the readings provided, and how they contribute to, and build on, knowledge within the discipline (Wade & Moje, 2001). Students need to be taught how to understand the purpose of the texts they are reading, identify the intended audience, and judge the credentials of the author and hence the validity of their claims within the text. Instructors can also talk about certain articles they provide to students, identifying why they value them, and why they have included them in the course readings. Since reflection is an important part of the learning process (Partridge, Edwards, Baker, & McAllister, 2008), instructors can offer opportunities for students to discuss and reflect on the readings together so they can begin to understand how knowledge is produced.

The Example: Source Justification Assessment

The Source Justification assessment created for a core first year course was a modified and extended annotated bibliography. The purpose of this assessment was to encourage students to look for sources early on, and to allow enough time for analysis and synthesis

of ideas as well as helping students understand how academics build on, and contribute to, knowledge development within the discipline.

The participating lecturer felt this assignment was important because of the high incidences of poor quality, apparently random, and mostly web-sourced references students had relied on for essays in previous classes. Most of the previous reference lists lacked any of the instructors' recommended readings, which were listed in the course outline, but which students had to access independently. This suggests that they either didn't look back at the course outline for readings, or that they weren't able to access specific sources, preferring instead to perform a simple Google search.

The Source Justification assessment required students to select five key sources relevant to their essay question, write the full APA reference and state how they had found the source. Then they were asked to identify five key points from the readings that supported the essay question, and justify why they felt this source was useful for answering the essay question, and how it connected to other sources they had found. Prior to completing the source justification, students attended a library session, and library tip-sheets were provided in class and later on-line. Once students had posted their assignments online, feedback on the source selection and justification was given one week before the essay submission.

Several things were learned from this task. Firstly, some students initially selected poor sources: a Wikipedia page and a Facebook fan page entry provide examples of poor selections students attempted to justify.¹⁴ The recommended readings were provided online for the first time (perhaps allowing easier access), so there was a stronger reliance on what the lecturer recommended. When identifying how they found the sources, few students made reference to selecting a source because they had found it through the reference list of another source, suggesting the connections between sources are lacking.

The initial aim was to pick only 20-30 of the most appropriate sources submitted to post online to provide a good model of appropriate sources. However, a technical difficulty meant all entries had to be approved, resulting in a resource of 150 sources, which became less useful than envisaged. Nevertheless, a review of all the reference lists indicated that one third of the students had references from the Source Justification online resource in their reference lists, and all ten focus group students said they had been online to look at what others were using.

Despite the setbacks, feedback from the focus group students about the assessment was positive. Students felt a key purpose of the assessment was to "start researching early", but also "to see how well we can critically evaluate sources and how we get information out of them" (Student focus group interview, personal communication, July 26, 2011). Other students said that being able to see what others had used helped them see if they were on the right track and provided interesting sources that they hadn't found.¹⁵ One student said they had not really ever given much thought to the sources they selected, just taking whatever they found, but the assessment encouraged reflection on how the source would be used.¹⁶

Overall, the impact of this assessment in getting students to think about source selection was positive. By having to identify the key points of the source, the task helped students

¹⁴ "Using a Facebook fan page as a source has certain advantages such as having the ability to be updated quicker than most sources and having close communication with other webpages such as beehive.govt.nz, so as to share relevant information from credible government sources" (Source Justification entry, May 5, 2011).

¹⁵ "I used it quite a lot for the essay, to see what kind of points I want to write about and see if any of the sources are on there that the other people had written about could answer what I needed" (Student focus group interview, personal communication, July 26, 2011).

¹⁶ "it got you thinking about what you are actually going to use out of the sources instead of just like looking at it and thinking oh yeah, this looks helpful. You actually had to think about yep I'm going to use it specifically for this and that" (Student focus group interview, personal communication, July 26, 2011).

clarify how the sources could contribute to their argument, and to be able to synthesise the ideas across texts more effectively. The task does require some modification, but the initial aims of getting students to start researching early and to carefully consider the relevance and use of the texts were successful.

Conclusion

There is no one solution to helping students become information literate and confident writers. However, actively engaging and encouraging students from first year to focus on the research process, and providing embedded information literacy and academic writing skills into content papers via formative assessments will help students succeed in their transition into academic literacy, and help them understand how knowledge is created within their discipline. As the gap between high school and university appears to be widening, and more students are entering university unprepared for the tertiary academic demands, it is essential for students to have explicit opportunities to develop academic information literacy skills.

An intentional focus on the research process will help students to understand that completing assessment successfully depends on having a robust research process. Of particular concern to lecturers participating in our research was a perceived inability of students to carefully select, evaluate and use information sources, particularly scholarly journal articles. Asking students to reflect on how they are conducting research may help them identify the challenges that may be hindering successful learning.

This paper has identified two activities that are being developed to improve information literacy through formative assessment in two first year courses. The interventions being trialled within the participating programme have had generally positive responses from both students and lecturers. The students have appreciated the efforts of instructors to help them become better researchers and better writers.

It is also essential that university instructors recognise their role in developing students' information literacy and writing skills, rather than relying on other university support services and on the students themselves seeking help should they need it. This change of perspective can be seen in all the participating lecturers through involvement in this research. One lecturer commented:

I hadn't consciously thought about it before becoming involved in the research and now I see that I have quite a clear responsibility in terms of [student] learning to teach them about information literacy, and how to be information literate, and how to actively incorporate that into the lectures. (Staff interview, personal communication, December 14, 2010)

This is the change we need to encourage more broadly if our students are to become information literate in a world of ever expanding information.

References

- Ahmed, R., & McMahon, K. (2006). The benefits of good writing. In S. Davies, D. Swinburne & G. Williams (Eds.), *Writing matters: The Royal Literary Fund report on student writing in higher education* (Ch. 1). London, UK: Royal Literary Fund.
- American Library Association Presidential Committee. (1989). *Presidential Committee on information literacy: Final report*. Retrieved from <http://www.ala.org/ala/mgrps/divs/acrl/publications/whitepapers/presidential.cfm>
- Andretta, S. (2006). Information literacy: The new pedagogy of the question"? In G. Walton & A. Pope. (Eds.), *Information literacy: Recognising the need*. Oxford: Chandos.

- Andretta, S., Pope, A., & Walton, G. (2008). Information literacy education in the UK: Reflections on perspectives and practical approaches of curricular integration. *Communications in Information Literacy*, 2(1), 36-51.
- Angier, C., & Palmer, W. (2006). Writing solutions. In S. Davies, D. Swinburne & G. Williams (Eds.), *Writing matters: The Royal Literary Fund report on student writing in higher education* (Ch. 3). London, UK: Royal Literary Fund.
- Barton, H. (n.d). *Information literacy: Learning how to learn*. Retrieved from http://www.ri.net/RITTI_Fellows/Barton/infolit.html
- Bhattacharya, M., & Jorgenson, L. (2008). Minimizing plagiarism by redesigning the learning environment and assessment. In T. Roberts (Ed.), *Student plagiarism in an online world: Problems and solutions*. NY: Information Science Reference.
- Brabazon, T. (2007). *The university of Google: Education in the (post) information age*. Hampshire, UK: Ashgate.
- Bundy, A. (Ed.). (2004). *Australian and New Zealand information literacy framework: Principles, standards and practice* (2nd ed.). Adelaide: Australian and New Zealand Institute for Information Literacy.
- Callahan, M. K., & Chumney, D. A. (2009). "Write like college": How remedial writing courses at a community college and a research university position "at-risk" students in the field of higher education. *Teachers College Record*, 111(7), 1619–1664.
- Carless, D., Joughin, G., & Liu, N. (2006). *How assessment supports learning: Learning oriented assessment in action*. Hong Kong: Hong Kong University Press.
- Chanock, K. (2001). *From mystery to mastery*. Retrieved from <http://learning.uow.edu.au/LAS2001/selected/chanock.pdf>
- Crozier, S. (2007). "Out damned spot": Removing the taint of remedial from learning development. In C. Fraser & L. Ayo (Eds.), *Anchoring our practice: Perspective, partnerships, projections: Refereed proceedings of the 2006 Association of Tertiary Learning Advisors of Aotearoa New Zealand Conference* (pp. 2-13). Auckland: ATLAANZ.
- Davis, B. G. (1999). *Motivating students*. Retrieved from <http://www2.honolulu.hawaii.edu/facdev/guidebk/teachtip/motiv.htm>
- Dolan, D. V., & Martorella, G. (2003). Discipline-based information literacy and the lifelong learner. *International Journal of Lifelong Learning*, 10, 1319-1334.
- Dreyfus, S. E., & Dreyfus, H. L. (1980). *A five stage model of the mental activities involved in directed skill acquisition*. Washington, DC: Storming Media. Retrieved from <http://www.dtic.mil/cgibin/GetTRDoc?AD=ADA084551&Location=U2&doc=GetTRDoc.pdf>.
- Emerson, L. E. (2008). Plagiarism, a Turnitin trial, and an experience of cultural disorientation. In C. Eisner & M. Vicinus (Eds.), *Originality, imitation and plagiarism: Teaching writing in the digital age*. USA: The University of Michigan Press.
- Emerson, L. E., Muirhead, J., & Stevens, S. M. (2008, November). *Scaffolding the writing process in a large first year science class*. Presentation at Symposium on Tertiary Assessment and Higher Education Student Outcomes, Wellington, New Zealand.
- Gaipa, M. (2004). Breaking into the conversation: How students can acquire authority for their writing. *Pedagogy Critical Approaches to Teaching Literature, Language, Culture and Composition*, 4(3), 419-437.
- Godwin, P. (2006). Keeping up with the Google generation: The challenge for information literacy teachers. In G. Walton & A. Pope (Eds.), *Information literacy: Recognising the need*. Oxford: Chandos.
- Grafstein, A. (2002). A discipline-based approach to information literacy. *Journal of Academic Librarianship*, 28(4), 197.
- Head, A. J., & Eisenberg, M. B. (2009a). *Finding context: What today's college students say about conducting research in the digital age* (Project Information Literacy Progress Report). Washington: University of Washington.
- Head, A. J., & Eisenberg, M. B. (2009b, December 1). *Lessons learned: How college students seek information in the digital age* (Project Information Literacy Progress Report). Washington: University of Washington.

- Head, A. J., & Eisenberg, M. B. (2010). *Truth be told: How college students evaluate and use information in the digital age* (Project Information Literacy Progress Report). Washington: University of Washington.
- Head, A. J., & Eisenberg, M. B. (2011, June 3). College students eager to learn but need help negotiating information overload. *The Seattle Times*. Retrieved from http://seattletimes.nwsources.com/html/opinion/2015227485_guest05head.html
- Hepworth, M. (2006). Reflections on the information literacy: Recognising the need conference. In G. Walton & A. Pope (Eds.), *Information literacy: Recognising the need*. Oxford: Chandos.
- Husain, S., & Waterfield, R. (2006). The first year of higher education. In S. Davies, D. Swinburne & G. Williams (Eds.), *Writing matters: The Royal Literary Fund report on student writing in higher education*. London, UK: Royal Literary Fund.
- Immerwahr, J. (2011). *Readings assignments*. Retrieved from <http://www.teachphilosophy101.org/Default.aspx?tabid=122>
- Jamieson, S., & Howard, R. M. (2011). *The Citation Project: Preventing plagiarism, teaching writing. Initial data from the Citation Project study of student use of sources in researched papers from 16 US colleges and universities*. Retrieved from http://citationproject.net/CitationProject-Sixteen_School_Study-Sources_Selected.pdf
- Lea, M., & Street, B. (1998). Student writing in higher education: An academic literacies approach. *Studies in Higher Education*, 23(2), 157-172.
- Leckie, G. J. (1996). Desperately seeking citations: Uncovering faculty assumptions about the undergraduate research process. *The Journal of Academic Librarianship*, May, 201-208.
- Lunsford, A. (2010). Writing and the profound revolution in access. *Project Information Literacy, "Smart Talks,"* 2, July 12.
- McGuinness, C. (2006). What faculty think: Exploring the barriers to information literacy development in undergraduate education. *The Journal of Academic Librarianship*, 32(6), 573-582.
- Middle States Commission on Higher Education. (2003). *Developing research and communication skills: Guidelines for information literacy in the curriculum*. Philadelphia, PA: Middle States Commission on Higher Education. Retrieved from <http://www.msche.org/publications/devskill050208135642.pdf>
- Murray, N., & Kirton, B. (2006). An analysis of the current situation. In S. Davies, D. Swinburne & G. Williams (Eds.), *Writing matters: The Royal Literary Fund report on student writing in higher education* (Ch. 2). London, UK: Royal Literary Fund.
- Partridge, H., Edwards, S., Baker, A., & McAllister, L. (2008). The reflective online searching skills (ROSS) environment: Embedding information literacy into student learning through an online environment. *ILFA Journal*, 34(1), 55-71.
- Pickford, R., & Brown, S. (2006). *Assessing skills and practice*. Abingdon, Oxon: Routledge.
- Pope, A., & Walton, G. (2006). *Information literacy: Recognising the need: An introduction*. In G. Walton & A. Pope (Eds.), *Information literacy: Recognising the need* (pp. 3-12). Oxford: Chandos.
- Proctor, L. (2006). Not such strange bedfellows! Supporting learning by embedding information literacy skills in academic assessment. In G. Grigg & C. Bond (Eds.), *Supporting learning in the 21st century: Refereed proceedings of the 2005 Association of Tertiary Learning Advisors of Aotearoa New Zealand Conference* (pp. 129-152). Auckland: ATLAANZ.
- Radloff, A. (2006). Supporting student learning in the 21st century: What's the job and whose job is it? In G. Grigg & C. Bond (Eds.), *Supporting learning in the 21st century: Refereed proceedings of the 2005 Association of Tertiary Learning Advisors of Aotearoa New Zealand Conference* (pp. 2-17). Auckland: ATLAANZ.
- Reynolds, P., Ker, H., & Day, B. (1996). Students' written skills in a New Zealand polytechnic. In L. Emerson, M. Franken, & G. Harris (Eds.), *The TWN Proceedings: Writing beyond the margins. Vol. 1*. Palmerston North: Massey University.

- University of Washington Libraries. (2010). *Personas development*. Retrieved from http://kwhitenton.com/libraries_personas/
- Wade, S. E., & Moje, E. B. (2001). The role of text in classroom learning: Beginning an online dialogue. *Reading Online*, 5(4). Retrieved October 1, 2011, from http://www.readingonline.org/articles/art_index.asp?HREF=/articles/handbook/wade/index.html
- Walden, K., & Peacock, A. (2006). The i-map: A process-centred response to plagiarism. *Assessment & Evaluation in Higher Education*, 31(2), 201-214.
- Walden, K., & Peacock, A. (2008). Economies of plagiarism: The i-map and ownership in information gathering. In C. Eisner & M. Vicinus (Eds.), *Originality, imitation and plagiarism: Teaching writing in the digital age*. USA: The University of Michigan Press.
- Wall, A. (2006). Introduction. In S. Davies, D. Swinburne, & G. Williams (Eds.), *Writing matters: The Royal Literary Fund report on student writing in higher education*. London, UK: Royal Literary Fund.

Reducing exclusions and improving attainment – how can that be? Evidence from a United Kingdom secondary school

Gwen Gilmore
Victoria University, Melbourne

Email: gwen.gilmore@vu.edu.au

Abstract

This paper presents discussion and exploration of the nature, extent and characteristics of a disciplinary Inclusion Room (IR). There is limited literature on systems at the interface of inclusion-exclusion. This research offers a unique theoretical lens on a contested domain of school discipline.

The research aimed to explore the nature, extent, and characteristics of the disciplinary inclusion situation. A mixed methodology included documentary analysis, a quantitative questionnaire of $n = 130$ students and $n = 30$ staff. Qualitative methods included document analysis and in-depth, semi-structured interviews with $n = 9$ students and $n = 9$ staff.

Students' and staff perspectives in this research illustrate how the IR strategy positively supported the goals of school improvement. Short periods of exclusion from the classroom to the IR ultimately supported school inclusion. Further, evidence suggests that reducing school suspension needs to be accompanied by a rigorous challenge to goals for student exclusion. Disciplinary inclusion is supported by the interplay of values and systems change, including achievement for all students, that supports reduction in suspension and inclusionary practice. Implications for NZ disciplinary provision include a critical examination of the nature and purposes of withdrawal rooms.

Keywords: *Inclusion, exclusion, discipline, student views on discipline.*

Introduction

This section introduces the research background and research questions. From 2005 to 2010 the author led a partnership of 25 United Kingdom (UK) schools, including three secondary schools. The purpose and goals of the partnership were improving school attainment, reducing school exclusion, and improving attendance. Each school met these goals by 2010 through significant reductions in fixed term (FT) exclusions and improving attainment. The schools served an overall population of nearly 10,000 students; the socio-economic profile for the area was such that it was amongst the poorest 20% in the UK. The research school achieved two awards in the *Times Education Supplement* awards presented to the top six UK secondary schools in 2010/11. In 2010, the school was judged as outstanding by the Office for Standards in Education, Children's Services and Skills, the official school inspection service.

The research question for the paper investigates the nature, extent, and characteristics of inclusion in this secondary school, from student and staff perspectives, in relation to the disciplinary inclusion room (IR). This paper presents summary outcomes of doctoral research, from students and staff, in one of these secondary schools. More detailed findings are reported in Gilmore (2010). The term 'FT exclusion' is used to indicate the UK process for student exclusion from school. The NZ equivalent is stand-down or suspension. The school makes a decision, on the basis of school rules, values or beliefs, that a temporary, school-imposed exclusion from school will serve a retributive function as described by Macfarlane and Margrain (2011). The purposes of FT exclusion are similar in both countries. The next section considers the literature and theoretical background for discipline.

Literature and Theoretical Background

The argument in this paper, that a disciplinary IR process can support inclusionary principles, captures the dialectic relationship between the production of key concepts of inclusion and professional practice within the school in relation to the IR. Hedegaard and Fleer (2008) note “this form of knowledge is based on the idea that knowledge generalisation is connected to the content of the problem areas that have produced the knowledge” (p. 10).

General Excellence in Cities (EiC) projects had been subject to overall evaluative reporting (Hallam, Castle, & Rogers, 2005), but there had been no reported or specific research on these forms of intervention. Rather, the UK press (Brettingham, 2007) reported on the provision from some parents’ perspectives. Further, it is acknowledged that many schools offer similar forms of student discipline either as ‘withdrawal’ or ‘isolation’ or other forms of discipline before an exclusion. The author was aware of other EiC partnerships that had similar goals for FT exclusion reductions and achievement gains, but the elements were not subject to research such as in this paper. The research therefore offered an opportunity to clarify the conditions within which this IR provision operated from staff and students’ perspectives and to clarify some points in a gap in research knowledge.

This research uses the definition of inclusion as a continuum of participation, equality and diversity concepts as articulated by Black-Hawkins, Florian, and Rouse (2007). Thus, I considered the nature of IR experience in relation to the principles of student participation and staff values for that participation in the school. In what ways did the school consider diversity in relation to the IR? Was the IR process an active and collaborative process based on mutual recognition and respect, or was it a final statement on a student discipline matter? In addition, was inclusion, in relation to the IR, framed in pedagogical terms identified by Skidmore (2004) as a schooling “discourse of inclusion”? (p. 113). Inclusionary education entails development of values within the school system and organisational structures.

UK literature is replete with examples and detail on the nature of exclusion and those who are more likely to be excluded (Parsons, 1999, 2009, 2011). At the time of the intervention it was those with Special Education needs (SEN) and certain ethnic minorities and boys who were most likely to be excluded in the UK. In NZ, the exclusion picture is similar with particular ethnic groups, Māori and Pasifika and particular students (especially boys) excluded in low decile schools (Ministry of Education, 2011). New Zealand does not collect data on the exclusion-suspension of pupils with SEN in the same way as the UK.

Very little research on school disciplinary approaches, such as the IR, is informed by students themselves. McKeon (2001) provides some research evidence for a similar discipline process but it combined instruction and counselling with disciplinary approaches. Further, where students’ perspectives on discipline were involved in provision there was a reliance on external approaches to the school, like pupil referral units (PRU). In addition, most research with students is completed with students who have been excluded (Hilton, 2006; Kane, 2006).

Traditional research into disciplinary provision, withdrawal rooms, detentions, contract rooms and the similarly titled secondary provision, aimed at supporting school discipline, has tended to be dominated and bounded by analysis illustrating systems of control. As Slee (1995) notes, the beneficiaries of these managerial approaches provide increased potential for “compliance, provide less disruption for the organisational tranquility of the school by removing pedagogy, curriculum and organisation from the field of therapeutic vision” (p. 68). Control remains myopic.

Research literature provides abundant evidence of ‘effective’ and ‘efficient’ behaviour management systems and data management approaches (McKeon, 2001). Pomerantz

(2007) considers how discourses of exclusion are bounded by considerations of control and procedural documentation. Ainscow (2007) suggests organisational change and better systems are necessary but not sufficient to change outcomes for those who are likely to be excluded. He further notes, somewhat ironically, that it was the behaviour and assumptions of the adults that needed to be examined. Fulcher (1989) identified the main discourse of theory and practice for inclusion in the UK as following an individualised and deficit model – if we fixed the child, we supposedly fixed the problem. Slee (1995) made similar observations in terms of professionals and experts, extending and intensifying control of students in situations of discipline that appeared to have little to do with education. Whose interests does school serve and for what purpose? Control and order might be one side of the schooling journey; student learning, rights to participation, respect for their diversity and success for all is another matter.

Qualitative research on student exclusion frequently reflects medicalised responses to solving, or otherwise, problems of the disaffected (Hilton, 2006; Jull, 2008). Neilson (2000) makes two linked points within this discourse of deficit: firstly, the negative influence of teacher beliefs and values on student learning potential; and secondly, she expands the deficit discussion in a closer examination of the 'medical model' for disability: the assumptions about what is 'normal' and what is not, and what needs to be fixed to make learning or behaviour 'normal'. Under medical regimes, professionals use their expert knowledge to solve problems, and fix the individual's deficits. Therapeutic, expert lenses and/or behavioural-orientated research on the excluded are the most usual research orientations (Fulcher, 1989).

Skidmore (2004) applies a wider systemic and organisational lens to examining inclusive practice which might provide opportunities for examining relationships amongst factors at different levels of analysis. In the current research perspectives of students and staff on inclusive policy, practice and culture are explored in relation to a facility, disciplinary IR, designed to reduce FT exclusion. Further, this IR research develops the Slee (1995) concern and challenge to "recover and reconstruct discipline as an educational concept" (p. 3) by asking the questions: How was it possible that the school had included more pupils, reducing both FT exclusion and improving attainment? Was the IR a mechanism for further control of their students? There were other possibilities, within the research, to examine an alternative hypothesis from Slee that the IR had an educational role.

Finally, in New Zealand the Freeman (2010) health report calculates that it is cheaper to retain students in school than to exclude them. Further, Freeman strongly suggests a consideration of the longer-term consequences for pupils. She indicates the puzzling nature of exclusionary processes and provides a case study from a student and parent perspective. The current overall Ministry of Education strategy for reducing school exclusion relies on a Positive Behaviour for Learning (PBfL) strategy (Ministry of Education, 2010) to be implemented in individual schools. There is some American evidence of success for this strategy (Bohanon et al., 2006; Horner et al., 2009). However, there is no current New Zealand evidence of sustained reductions in school suspension and stand-down using this approach.

A NZ evaluative report on restorative approaches to reducing school suspension (Gordon, in press) shows a 60.5% reduction across 10 schools that developed restorative justice approaches to discipline in comparison with all NZ primary and secondary schools. Nevertheless, the Gordon report indicates that restorative justice has not yet illustrated ongoing and sustained improvements in engagement in learning. Issues like early leaving of particular groups, such as Maori, is still a problem in some schools.

In summary, there may be value in examining an approach to reducing school suspension from the UK that developed a clear strategy for inclusionary practice within a learning environment that challenged the value and efficacy of student exclusion.

Disciplinary Inclusion Room: What is it?

The secondary school in this research was allocated funds from the EiC Partnership to develop its internal IR provision, reduce FT exclusions, and improve attainment. This school spent approximately 20% (£95,000) on directly staffing the IR. The balance of this fund supported other pastoral staff like Learning Mentors (LM), Learning Support Units and Gifted and Talented (G & T) staffing.

The IR room was developed in three phases between 2005 and 2009. Firstly, the partnership leaders worked to national guidance documents with the head teacher whilst setting overarching goals for reduction of FT exclusion. Secondly, the SEN Coordinator and other leadership staff reviewed regional IR provision elsewhere with similar goals to reduce FT exclusion. Thirdly, the school consulted the rest of the staff, students and governors to develop the ethos and principles around the programme. This was an ongoing process and the systems were reviewed each year with staff and on a continuing basis with individual student feedback.

The IR physical space was uninviting and unattractive, a judgement supported by the students and staff in this research. This was an intentional dimension to the provision. The students sat in individualised booths with little stimulus other than the rules and curriculum lessons to complete. The day started at 12pm and finished at 5pm with one break. This meant that students arrived when other students were in class. Their only break was when the rest of the students were leaving school.

Students attended the disciplinary IR for two reasons: firstly, for disregarding school detention systems; and secondly, for matters of a more serious disciplinary nature where they previously would have received an FT exclusion. The issue of an FT exclusion is largely contextual. Students had initial conversations about the incident, a written statement at the end of the IR time on what they were going to do in future, and some post conversations with parents and senior staff for those involved in more serious incidents such as bullying or racism, for example. Student participation in the process was relatively passive although they were listened to and could have a point of view. Nevertheless, the IR retained the students within the school responsibility and appeared to provide space for self-reflection.

The formal rules for the operation of IR were clearly set by the adult staff in this school: the 12pm to 5pm regime, parental involvement at the beginning and end, silent learning linked to their curriculum environment, and two escorted breaks. For those who had committed more serious offences, for example assault or bullying, there was ongoing surveillance in the form of additional reporting. The student and staff interviews reflected these specific rules.

The rest of the staff, for their part, considered the IR manager to be doing a good job in maintaining the inclusionary dimensions of the role, sourcing appropriate work, and identifying problems with particular students. For example, she enabled senior staff to identify particular teachers who were not differentiating work for students with SEN on homework tasks. Her analysis each term enabled the staff to be more rigorous on issues of fairness, pastoral support and other inclusionary school functions.

Data analysis between 2008 and 2010 on the use of this IR room confirmed that the room was still used, on a gradually reduced basis, and for which an FT exclusion might have been given in other schools. Analysis of the student profiles of those who attended the IR revealed stark contrasts with national data sets. In this school, students with SEN were no more likely to go into this IR, unlike national UK statistics which indicate they are eight times more likely to be excluded (Department for Education, 2011). In addition, the proportion of students who attended IR and who were on Free School Meals (FSM) were no more likely to go to the IR than other students. Again, this was unlike national statistical datasets that showed students with FSM were three times more likely to be excluded

(Department for Education, 2011). These data suggested that the IR was being used relatively equitably compared with national UK data for student exclusion. It also suggested elements of fairness not reflected in national exclusion data. I elaborate on points of fairness and equitable use of provision in the research findings.

Further data analysis showed students were unlikely to be in the room for more than three days. If they had a disciplinary IR more than three times, or the reason for the intervention was judged to require additional pastoral support, then personalised alternatives were developed that maintained the students within the school system. The next section considers the literature on exclusion and inclusion.

Methodological Background, Ethics and Methods

This section explains the methodological assumptions, school documentation evidence, the research participants, ethical procedures and methods.

The research methodology was framed by concepts for Cultural Historical Activity Theory (CHAT). Hedegaard and Fler (2008) indicate that CHAT provides an opportunity to focus on the perspectives of the individual and the institution's construction of the activity. This research considers the concepts of inclusion as an activity system mediated by the students and staff in relation to the IR. Conflict and motives include the nature, extent, and characteristics of inclusion in relation to the disciplinary IR. Within a CHAT framework questions such as "What was the nature of the conflict?" "What were the characteristics, power and community of the staff motives for students going to the IR?" and "What, in turn, were the students' perceptions of those rules?" were examined.

Within CHAT rules are the explicit and implicit regulations, norms and conventions that constrain actions and interactions (University of Helsinki, 2009). The activity system is bounded by the IR process. Thus, this research considers the nature and characteristics of the students, their talk about why they were in the IR, and how they talked about the IR in relation to the talk of the staff. Community within CHAT considers those individuals and subgroups who share the same general purpose. Students examine who are helpful, or not, amongst their peers to support them to not go to the IR. The nature and characteristics of the IR were triangulated between the students and staff community including dilemmas and contradictions within those communities. The division of labour and power reflect the horizontal division of tasks between members of the community and the vertical division of power and status. Thus, this research enabled an examination of the nature of the students' power in relation to entry and exit from the IR, and who undertook the various roles and to what purpose or outcomes.

CHAT was selected in order to clarify the discourse rules, power and community of students and staff engaged in the disciplinary IR situation. This methodology was chosen in order to overcome the more traditional individualistic and behavioural approaches to examining discipline practice and policy. The research focused on examining the qualities of the social interactions, the language, the policy and motives of those engaged in this disciplinary IR process.

Inclusionary principles in this paper – participation, diversity and equality – were contained within the Booth and Ainscow (2002) *Index for Inclusion* and were used to administer an on-line questionnaire with Year 8 and 9 students. These groups were selected as they were the most and least likely to receive a disciplinary IR, respectively. One hundred and sixty (47%) completed this questionnaire (Table 3 below). Nine Year 8 & 9 students (9%), who attended the IR during autumn 2008, also completed the questionnaire (Table 4). The purpose of the different administration was to determine any quantitative differences amongst students who had undertaken an IR process compared with the more general student population.

Questions were adapted from Booth and Ainscow's (2002) *Index for Inclusion*. These data were gathered using an on-line questionnaire to Year 8 and 9 students during two weeks of spring 2009. The purpose of this questionnaire was to gather background contextual information about inclusionary concepts while specific conclusions were gathered at a broad level. This information is summarised in Table 3.

A purposive group of nine students who attended the IR during autumn 2008 also completed the same questionnaire by hand, representing a 10% sample of those who attended the disciplinary IR (see Table 4). The purpose of a separate analysis was to gain an understanding of any differences in perceptions between an overall Year 8 and 9 group and these students.

Further semi-structured interviews were conducted with these nine students. These interviews followed aspects of the on-line questionnaire with specific questions to clarify the nature of student understanding of participation, diversity and equality. Further detail, the on-line staff questionnaire and staff interviews, are reported elsewhere in Gilmore (2010).

Sam, Ronaldo and James were students who took part in the research as students who attended the IR (all names are pseudonyms). The reasons they attended the IR were related to detention or relatively minor disciplinary matters. Callum, Leny, Lewis, Peter, Chris and Bob were research students from the group who would previously have received a FT exclusion. They attended the IR for more serious reasons for which in 2005 students would have received a FT exclusion such as fighting or assault. Four of these research students were involved in an assault, endangering others and fighting. The incidents these research students were involved in were not in classrooms. Leny and Peter were involved in what the school considered computer-related incidents.

Ethical procedures followed the University of Exeter protocols as well as the local authority ethical procedures. These included informed consent from each student, their parents, and the staff. Students were included in a discussion on the meaning of informed consent and their right to withdraw at any time.

My leadership role, across the partnership, had a two-fold lens. One was to support individual schools. In-school roles included professional development for pastoral roles such as staff managing the disciplinary IR, and LM. The second, wider partnership role, was managing educational staff for G & T and a Learning to Learn strategy (Campaign for Learning, 2011). New Zealand (NZ) educational experiences, firstly facilitating a professional development contract with the Christchurch College of Education, and secondly with Special Education 2000 in 1999-2000 using inclusionary principles, were important to informing my values for this UK project. Potential ethical dilemmas between the leadership role and research roles were managed through supervision, data checking by participants, and close attention to ethical principles. I used the principles and guidance from Cohen, Manion, and Morrison (2000) particularly in ensuring boundaries were maintained around my researcher role and my professional role in the school. Supervision and participant checking of initial transcripts informed the ethical dimensions in the research. Students and staff chose their own pseudonyms. Further ethical details are contained in Gilmore (2010).

Methods reported in this paper include document analysis, quantitative student and staff survey on inclusionary principles, and aspects of the student and staff interviews. Documents examined included the school website, professional note-taking during the two years of the project, LA data, internal school records of IR, and national data on inclusion rooms. This information is threaded through the discussion below.

Results

(a) Document Analysis

Document analysis included examining data on actual achievement and FT reduction within the school and partnership schools between 2006 and 2010. Table 1 illustrates 2006-10 General Certificate for Secondary Education (GCSE) results.

Table 1: Attainment for the research school, partnership and local authority (LA)

	2006	2007	2008	2009	2010
Research school 5 GCSE A* - C	49%	64.6%	68.6%	72%	68.2%
<i>Three schools-partnership average</i>	<i>42.4%</i>	<i>55.1%</i>	<i>55.9%</i>	<i>58.6%</i>	<i>63.5%</i>
<i>Local Authority average (LA)</i>	<i>57.3%</i>	<i>58.5%</i>	<i>61.4%</i>	<i>68.7%</i>	<i>72.3%</i>
National average	57.5%	60.1%	64.5%	69.8%	74.7%

GCSE is a similar qualification to the New Zealand Certificate of Educational Achievement (NZCEA) level 1 qualification results. (Data updated and provided by LA, March, 2011). Five grades between A*-C were considered satisfactory school exit standard.

Student achievement in the research school moved from 49% to 68.2% achieving a GCSE grade. This was a significant improvement in learner outcomes. To illustrate that these improvements were not simply confined to one school, the partnership schools' average improved from 42.4% to 63.5%. The school and the partnership schools were initially well below LA and national averages. This was a slightly greater improvement than LA (15% gain) and national average gains (17%). At the start of the research, the research school was below LA average but above by 2010. In the UK an annual shift of 2% in attainment was regarded as an achievable expectation (Crowe, 2008). It was evident that there were significant improvements annually across this school as well as the three partnership schools that used similar approaches to discipline and achievement.

Table 2: Fixed Term Exclusions 2005-2010

	2005/06		2006/07		2007/08		2008/09		2009/10	
	No of incidents	n/1000	No of incidents	n/1000	No of incidents	n/100*	No of incidents	n/100*	No of incidents	n/100*
Fixed term exclusion the research school	82	105.5	12	12.9	4	0.5	4	0.5	6	0.7
Total for partnership	229	64.3	153	43	90	2.6	17	0.5	21	0.6
Total for Local Authority	2538	7.5	1895	4.1	1853	0.7	1699	0.3	1311	4.1

* In 2007/08 the indicator changed from a per 1000 pupil ratio to a per 100 pupil ratio. Percentages are rounded. Data taken from LA records

Table 2 shows the potential for a young person to receive a fixed term exclusion being reduced over time in this school. In 2005, the start of the disciplinary IR programme, a pupil had a 10% chance of receiving a FT exclusion. During the following four years the figure halved each consecutive year. For 2010, this is now at a 0.1 level, or less than 0.05 percent chance of getting a FT exclusion. The table shows the school to be significantly below LA and National exclusion data from the 2005-06 figure, which was double LA levels.

(b) Questionnaire

The questionnaire was administered as an on-line survey to all Year 8 and 9 students during a two-week period in January, 2009.

Table 3: Summary of 47% Year 8 and 9 Student Questionnaire. n = 160

(n = 160)	Year 8 N = 95 (65%) total possible		Year 9 N = 65 (30% total possible)	
	Definitely or mostly agree	A few times or disagree	Definitely or mostly agree	A few times or disagree
1. Teachers try to make the lessons easy to understand for me	71	24	39	26
2. We are taught to appreciate other people who have different backgrounds from our own.	79	16	52	13
3. I usually know what will be taught in the next lesson	48	47	27	38
4. In my lessons teachers usually expect students to help each other	61	34	31	34
5. In most lessons students behave well towards each other	75	30	25	40
6. In most lessons teachers behave well towards me	81	14	43	22
7. Teachers help everyone who has difficulties with lessons	71	24	35	35
8. Teaching assistants work with anyone who needs help	71	24	31	34
9. Homework helps with learning and is properly explained	48	47	12	53
10. Activities outside lessons interest me and others in my group	63	32	31	34
11. This school makes it easy for any children to attend	81	12	44	20
12. When I first joined the school I was helped to settle and feel part of the school	77	18	42	22
13. Teachers seem to like teaching us	63	31	31	33
14. Teachers and other staff 'sort out' difficulties with behaviour without wanting us to leave the school	76	19	35	29
15. Teachers work hard to make this school a good place to come to	70	24	43	21
16. Everything possible is done to stop bullying	71	24	37	27
17. All children are made to feel welcome in this school	84	11	48	17
18. Students help each other	75	20	36	29
19. Staff help each other	72	22	46	19
20. Students and staff treat each other with respect	70	25	34	31
21. Teachers do not favour one group of children over others	59	36	28	37
22. Teachers try to help all us students to do their best	76	19	46	19
23. Teachers think all students are equally important.	64	31	33	32

Code. Practice = 1-10, Policy = 11-15, Culture = 16-23

(The across and down numbers are not always equal to the total number as some students were able to skip some questions)

The results illustrate that 47% of students in Year 8 and 9 who completed the questionnaire were generally positive about practice for inclusion in this school. Particularly relevant to the concept of equality, a majority of students perceived the school to appreciate students of backgrounds different to their own. Other question examples illustrating participation included being made to feel welcome in the school, teachers helping students to do their best, and perceptions of staff helping each other. Most students in Year 8/9 identified that they were helped to become part of the school.

The questionnaire indicated less spread of agreement by Year 9 students in terms of homework (Q.9), which they indicated did not contribute to learning. In response to knowledge of forthcoming lessons (Q.3), Year 9 clearly indicated that this is generally known and therefore they were less likely to indicate that students help each other in lessons (Q.4). The overall impression from Year 9 students was that they perceived the policies in the school to be providing at least some aspects for an inclusive policy with

Year 9 being less positive than Year 8 on some dimensions of the questionnaire. The Year 9 sample was smaller and may well have been answered by students who were more dissatisfied than those of a larger sample. Nevertheless, the difference with the Year 9 group is indicative of the need to treat such questionnaire data sensitively and perhaps survey all year groups when examining inclusion.

Table 4 illustrates the perspectives of a purposive student sample, selected from those who were part of the disciplinary IR procedure in autumn 2008. Therefore, these students could potentially have been part of a school exclusionary process in a previous school system.

Table 4: Summary of Inclusion Room Student Questionnaire. n = 9

(9% population using the IR during Autumn 2008)	Inclusion room Group: Year 8. N = 4		Inclusion room group Year 9. N = 5	
	Definitely agree/ agree	Disagree in a few lessons or disagree	Definitely agree/ Agree	Disagree in a few lessons or disagree
1. Teachers try to make the lessons easy to understand for me	2	2	5	0
2. We are taught to appreciate other people who have different backgrounds from our own.	2	2	5	0
3. I usually know what will be taught in the next lesson	3	1	4	1
4. In my lessons teachers usually expect students to help each other	2	2	5	0
5. In most lessons students behave well towards each other	3	1	4	1
6. In most lessons teachers behave well towards me	2	2	4	1
7. Teachers help EVERYONE who has difficulties with lessons	2	2	5	0
8. Teaching assistants work with anyone who needs help	3	1	5	0
9. Homework helps with learning and is properly explained	2	2	5	0
10. Activities outside lessons interest me and others in my group	3	1	5	0
11. This school makes it easy for any children to attend	3	1	4	1
12. When I first joined the school I was helped to settle and feel part of the school	3	1	5	0
13. Teachers seem to like teaching us	2	2	4	1
14. Teachers and other staff 'sort out' difficulties with behaviour without wanting us to leave the school	3	1	2	3
15. Teachers work hard to make this school a good place to come to	3	1	4	1
16. Everything possible is done to stop bullying	3	1	4	1
17. All children are made to feel welcome in this school	3	1	5	0
18. Students help each other	3	1	4	1
19. Staff help each other	2	2	5	0
20. Students and Staff treat each other with respect	2	2	4	1
21. Teachers do not favour one group of children over others	2	2	3	2
22. Teachers try to help all us students to do their best	3	1	4	1
23. Teachers think all students are equally important.	1	3	4	1

Results from the above data indicate the IR group was generally positive about inclusive practice, although this table illustrates some differences amongst Year 8 and 9 in the IR research group. However, the numbers are too small to draw firm conclusions. The information also illustrates that this smaller group of students could be described as being at least as positive, in the majority of dimensions, as the rest of the students in Year 8/9. For example, on inclusive teaching practice (Q.4), the perception that teachers expect students to share learning, this IR group of students was more positive than the overall cohort 7:2 as opposed to 9:7 ratio in the overall cohort. The Year 8 students had a similar perspective on the (lack of) value of homework for their learning as did the rest of the Year 8 cohort. This IR group was more positive than the overall group, which was more negative about the value of homework. Their thinking and feelings on the nature of

inclusive practice in the IR, and the rest of school time, were developed in the more extensive interviews below. Nevertheless, it might have been expected that these students would be more dissatisfied and potentially disaffected on a questionnaire of this nature.

Staff Perceptions

Thirty staff (30%) completed a similar questionnaire. It is reported in further detail in Gilmore (2010). Most staff who replied to the questionnaire definitely agreed, or agreed to some extent, in the majority of inclusion dimensions, policy, practice and culture. Of a total of 43 dimensions in the Inclusion Index, only three items had less than a third of staff disagreement. Of these three, only one (Q.11), the dimension on extended services, had a less than 50 percent positive response.

In summary, questionnaires of this nature are almost inherently potentially open to differing interpretations by respondents (Cohen et al., 2000). For example, the statement “behave well towards other students” is very open to interpretation. We can see from both data tables that the overall Year 8 and 9 students and cohort IR students in the school were similar for perceptions of the school policy, practice and culture for inclusion. Staff were similarly positive about the inclusive dimensions of the questionnaire. However, interviewing clarified the student and staff perspectives of the activity of the disciplinary IR.

Discussion

Students remain in this school for disciplinary, educational, and inclusionary purposes. They stay with their peers and within the lens of the school as an educational institution. Many of these incidents would result in FT exclusion in other schools. Staff strongly reinforce the message that if you exclude students, then learning is not occurring.

It was evident from student and staff interviews that sending students home was not a punishment or an effective disciplinary measure. Students and staff gave a number of reasons, including parents not having control of teenagers, parents’ attitudes to school and learning, and parents rewarding students instead of the process being a punishment. Lotte, as a Head of Year 8, observed that the differences in discipline have hugely improved compared with the previous three years. This was evident in the nature of the behavioural incidents she was managing; they were now fewer and at a lower level of intensity compared to 2005/06. There was also a very strong sense from her that the IR had an educational benefit. Most importantly, disciplinary IR maintained students’ participation in the school and their learning.

Sam, a student who received an IR discipline, reflected that his parents were disappointed in him “getting an IR”. Nevertheless, from his perspective interviews indicated this school IR control was considered proportionate and sufficiently negotiated by him so as to not alienate him further from the educational purposes within the school. The staff, in turn, recognised student wishes and used their discretion in enacting procedures for the IR. Staff interviews generally illustrated professional judgement and discretion, a participative and collaborative process with students, as proposed by Black-Hawkins et al. (2007).

Participation in the IR is a double-edged sword. Retaining students within the school, alongside student and teacher satisfaction, proved a surprising research finding. Students indicated the ‘in-school’ punishment system kept them within the school structure and system. However, they had different perspectives to their participation. Chris and Bob were two students who had experienced FT exclusion experiences in other schools. Chris, excluded from both his primary school and the PRU¹ noted the IR experience was “over the top” and that detentions were adequate. However, I noted that Chris remained in this

¹ Pupil referral units (PRU) were similar to Alternative Education provision in New Zealand. It was rare, but not unknown, that primary pupils were excluded from school on a semi-permanent basis

school, gaining his GCSE in 2011. Bob, who had a FT exclusion in another secondary school, indicated this disciplinary IR and the staff gave him “a second chance”. In other words, these forms of disciplinary provision, to be fully inclusionary, require both inclusionary staff relationships and provision of supports to increase the students’ sense of responsibility.

Pastoral matters were managed in a separate, although integrated, system through cross-checking IR data with a range of pastoral and professional staff. Staff interviews illustrated that the IR maintained students within the school processes and systems. Chris, for example, noted that IR was better than being sent home although it was challenging. Bob indicated that the practice of retaining the students in the school through the IR was an effective punishment that still retained a sense of belonging and participation. Further, he remained in the school before going to College in 2010.

The lack of ability to work with peers and having to work in silence were the most disagreeable aspects of the IR for all students. Both Chris and Ronaldo referred to it as being difficult. Peter and Callum both mentioned the booths and the way you had to put your hand up and ask questions as being additional layers of discipline. Clearly, this was the point from the staff’s perspective that students would find something disagreeable about the experience.

In terms of equality of opportunity, staff reflected that some students enjoyed the quiet working environment. One student in particular started to manipulate his behaviour in order to be put there. Staff changed the disciplinary response within the school. However, if the school was to be fully participatory, then staff had to consider how this student, and possibly others, could get some quiet learning time. Ronaldo and Lewis both noted that the IR gave them a chance to see how much work they could get done without the distractions of others and in this sense gave them a chance to fully participate back in regular classrooms. Lewis, in particular, appeared to recognise that this thinking time had been valuable to him considering his future.

Formal rules and the professional practice of the IR manager meant that students were provided with lesson material from staff. This replicated legislative guidance (Teachernet, 2010) for students being sent home to continue with a school-based programme. Students generally considered replication in the IR useful, in the sense that it kept them in contact with the lessons and school. Staff also indicated that the maintenance of lesson material was useful for them, although some teachers, those with languages, stated the connection with their subjects was difficult. Paul indicated how he took additional care, scaffolding lesson material for example to re-engage the student in the classroom, when a student had been in the IR.

Bob’s case highlighted the effectiveness of the IR intervention. The formality of the IR experience, combined with the focused attention of Alan (the Deputy Principal), Graham (a pastoral manager) and Lotte, seemed to have contributed to Bob’s relatively positive experience with the IR and the school in general. Although he had been part of an IR in his previous school, a combination of the expectations from his managed move and the informal attention of Graham meant he had a relatively successful transition into this school during the research period. Bob noted he found the rules on behaviour different and it took some time for him to adjust to a new school system. Graham’s interview revealed he had put in place a number of informal mechanisms for pre-empting problems with some young people. Graham had made informal contact with Bob after his IR visit, monitoring him each day through chats and conversations. He set up a ‘time out’ card for him, which Bob later indicated had supported him. As Artiles and Dyson (2005, p. 50) suggest, the “best interests of the student” are being considered here by staff in this school. Goodson and Hargreaves (1996) refer to this phenomenon as “the collegial gaze” (p. 120) – joint work to solving problems with students, linking the discretionary educational functions with inclusionary principles.

Chris provides another perspective on the activity of the IR for students with SEN. In NZ this support might have been similar to Resource Teacher of Learning and Behaviour (RTLb) intervention. The interview with the Deputy Head uncovered a contradictory narrative that illustrated a subtle and informal lens on this school's application of inclusionary principles. The school's judgement was that Chris needed to conform to the IR experience as a punishment. Alan indicated he had had lengthy conversations with Chris's parents about his attendance at the IR. Alan made it clear a SEN diagnosis did not provide an entitlement for preferential treatment. The school was known to use discretion and exceptions about IR, but still maintained the principle of inclusion. In the case of Chris, the educational nature of that judgement seemed to have been upheld.

The IR forms a division of labour/power bases that is part of the framework for social justice within this school. As is mentioned above, the IR experience appeared to be sufficient for students to understand the school discipline culture. Skidmore (2004) identifies the importance of examining power in issues of inclusion as they support understanding of culture, perhaps as a 'bridge' between agency and structure. The issue under examination then turns 'to what extent do students and staff exercise their agency and power?'

The IR experience appeared to be sufficient for students to understand the school discipline culture. Students in this research clearly use their agency albeit in particular ways. Staff exercised their considerable discretionary power and authority, generally 'with' colleagues and 'with' students; Bob and Graham, for example. At the same time students recognised their actions have consequences. Students felt the reasons for their IR experiences were "fair".

Key Messages

This paper considers the forms of assumptions the students and adults made in a situation where pressure for exclusion reduced. These findings provided a challenge to this research school's situation and posed the question: Was the IR simply a further control mechanism for the school or did educational and inclusionary outcomes for students improve?

There were four key messages from this study which linked to existing literature. They are:

- Sending students home was not a punishment, nor was it educational. Both students and staff indicate disciplinary IR was proportionate to their disciplinary event, and a negotiated approach that emphasised belonging. This school has gone some way to addressing the challenge of Slee (1995) to recover and reconstruct discipline as an educational and strong pedagogical construct.
- Staff and students agree that participation in IR was an ambiguous strategy in terms of student control. On the one hand students are temporarily excluded from contact with peers whilst "being there" (Black-Hawkins et al., 2007). On the other hand students indicated they had a sufficiently active and collaborative role that enabled them to remain within the educational realm of the school.
- The IR reinforces the complementary, educational discourse on school values and culture for inclusion and achievement.
- The IR forms part of an overall framework that all students can succeed and belong to the school. This position follows Parsons' (1999) principles of social justice and equity.

This school, in particular, showed that it was possible for disciplinary systems to promote inclusion. I argue that this transformation occurred as a result of a collective and eclectic process whereby the IR rules, power, and community intersect with and complement the overall purposes of the school.

The development of this IR provision was accompanied by a clear focus on the educational purpose for that provision and that educational purposes were directly linked to overall school goals for improving achievement. This disciplinary IR succeeded in relation to the overall purpose because the systems and organisational changes were perceived by staff and students to be fair. The students recognised the need to have a system of reminders about the school rules. The students still participated in the life of the school. The success of this intervention in this school must be viewed in context. The inclusive disciplinary system above was negotiated with appropriate staff input and student judgements on the particular disciplinary matter. The staff continually talked about the reasons why students were in IR and the pastoral systems needed to support the students, both in terms of curriculum adaptations and seeking external supports of multi-agency or LA partners. In addition, staff were challenged on their assumptions about students, and their classroom management. In one example a student received an IR for a matter the manager considered inappropriate and action was taken against that staff member. Equally, staff continually talked about and acted on their learning goals for the students which were informed by assessment for learning and Learning to Learn approaches.

Findings from this work are able to be generalised as the research develops experience of the school in a naturalistic manner and is illustrative of expectations rather than formal predictions. Nevertheless, schools can use the findings to consider how a disciplinary IR can complement educational processes through increasing participation, equality and diversity. Goals for inclusion can be enhanced through collaborative partnerships and active, ongoing engagement amongst students and staff to develop the educational experience.

Reflections for New Zealand

Tools like the *Index for Inclusion* (Booth & Ainscow, 2002), as used in this research, could support systems and organisational change and thinking about the overall purpose of the current PBfL strategy (Ministry of Education, 2009). This research, and experience with the partnership schools, suggests many of the systems set up in schools for withdrawal send students a mixed pastoral and disciplinary message. In such circumstances both groups take advantage of the situation and student and staff receive mixed educational messages. The research school understood this, and adjusted discipline, pastoral and pedagogical leadership. They carefully selected the staff for the provision to ensure continuity, presence, and to provide a consistent message about IR discipline. The periods of student stay in the IR were not lengthy as the partnership schools did not find sending students to the IR for longer periods of time more effective. They thought about who was in the IR, when and why. Parents were fully informed and involved at all times. Attendance at IR was a serious educational matter. Pastoral roles were included, formally and informally. Clearly, further NZ-based research is needed, and the findings from this research trialled as a preventative model to suspension and stand-down.

Withdrawal, contract rooms or other provision that might be associated with final steps before suspension require further attention and examination within wider school values for inclusion. Inclusionary practice should be informed by acknowledging diversity, participation and educational values. This research information illustrates the limitation on a singular focus on behaviour data. Such a focus is inevitably negative and sets staff and students against each other if schools are not clear about their educational remit. On the one hand, you get what you measure and talk about. On the other hand, behavioural data used with sensitivity and a clear purpose inform wider inclusionary and achievement goals as shown in this study.

In summary, this research suggests that disciplinary inclusion, reducing suspensions and improving attainment is dependent on organisational and systems development, including considering the role of RTLB in systems change that examines educational values and inclusive principles. Disciplinary approaches, detentions or withdrawal rooms should be matched with clear goals for achievement, ambition, and progress for all students. This requires strong pedagogical leadership, action on curriculum, pedagogy, and pastoral systems which are also ambitious for all our children and young people.

References

- Ainscow, M. (2007). Taking an inclusive turn. *Journal of Research in Special Educational Needs*, 7(1), 3-7.
- Artiles, A., & Dyson, A. (2005). Inclusive education in the globalization age. In D. Mitchell (Ed.), *Contextualizing inclusive education: Evaluating old and new international perspectives*. New York: Routledge.
- Black-Hawkins, K., Florian, L., & Rouse, M. (2007). *Achievement and inclusion in schools*. London: Routledge.
- Bohanon, H., Fenning, P., Carney, K. L., Minnis-Kim, M. J., Anderson-Harriss, S., Moroz, K. B., . . . Pigott, T. (2006). Schoolwide application of positive behavior support in an urban high school: A case study. *Journal of Positive Behavior Interventions*, 8, 131-145.
- Booth, T., & Ainscow, M. (2002). *Index for inclusion* (2nd ed.). Bristol, UK: Centre for Studies on Inclusive Education (CSIE).
- Brettingham, M. (2007, June 15). Parents in uproar over isolation units, *Times Education Supplement*, p. 11.
- Campaign for Learning. (2011). Phase 4 case study Retrieved September 20, 2011, from http://www.campaign-for-learning.org.uk/cfl/learninginschools/projects/learningtolearn/l2l_in_schools_case_studies.asp
- Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th ed.). London: Routledge/Falmer.
- Department for Education. (2011). Exclusion data. Retrieved October 25, 2011, from <http://www.education.gov.uk/cgi-bin/rsgateway/search.pl?keyw=058&q2=Search>
- Freeman, N. (2010). *When one door closes: Evidence based solutions to improve outcomes and open new doors for students excluded or expelled from school in New Zealand*. Wellington: Department for Health. Retrieved from http://cst.org.nz/groups/media_alerts/messages/topic/18n4bs00tAE8gDp7xRVbBG
- Fulcher, G. (1989). *Disabling policies? A comparative approach to education policy and disability*. London: The Falmer Press.
- Gilmore, G. (2010). *Inclusion and professionalism: Reducing fixed term exclusions in a South West secondary school* (Unpublished educational doctorate). University of Exeter, UK. Retrieved from <https://eric.exeter.ac.uk/repository/handle/10036/108628>
- Goodson, I., & Hargreaves, A. (Eds.). (1996). *Teachers' professional lives*. Abingdon, UK: RoutledgeFalmer.
- Gordon, L. (in press). *A preliminary evaluation of restorative practices in New Zealand schools: Ten case studies*. Networkers. Evaluation, training and research.
- Hallam, S., Castle, F., & Rogers, L. (2005). Research and Evaluation of the Behaviour Improvement Programme (pp. 158). London: Institute of Education, University of London.
- Hedegaard, M. (2008). The cultural historical theory of children's development. In M. Hedegaard & M. Fler (Eds.), *Studying children: A cultural historical approach*. New York: Open University Press.
- Hedegaard, M., & Fler, M. (2008). *Studying children: A cultural historical approach*. Maidenhead, UK: Open University Press.
- Hilton, Z. (2006). Disaffection and school exclusion: Why are inclusion policies still not working in Scotland? *Research Papers in Education*, 21(3), 295-314.
- Horner, R. H., Todd, A. W., Lewis-Palmer, T., Irvin, L. K., Sugai, G., & Boland, J. B. (2009). The school-wide evaluation tool (SET): A research instrument for assessing school-wide positive behavior support. *Journal of Positive Behavior Interventions*, 8, 10-23.
- Jull, S. (2008). Emotional and behavioural difficulties (EBD): The special educational need justifying exclusion. *Journal of Research in Special Educational Needs*, 8(1), 13-18.
- Kane, J. (2006). School exclusions and masculine, working-class identities. *Gender and Education*, 18(6), 673-685.
- Macfarlane, A., & Margrain, V. (2011). He Tapuwae o Mua: Footsteps towards responsive engagement with challenging behaviour. In V. Margrain & A. Macfarlane (Eds.),

- Responsive pedagogy. Engaging restoratively with challenging behaviour.* Wellington: NZCER Press.
- McKeon, M. (2001). Promoting the inclusion of students at risk of exclusion. *Emotional and Behavioural Difficulties*, 6(4), 236-250.
- Ministry of Education. (2009). *Positive behaviour for learning: Action plan, 2010-2012.* Wellington, New Zealand: Author.
- Ministry of Education. (2010). *Positive behaviour for learning.* Retrieved December 9, 2011, from <http://www.minedu.govt.nz/theMinistry/EducationInitiatives/PositiveBehaviourForLearning.aspx>
- Ministry of Education. (2011). *Stand-downs, suspensions, exclusions and expulsions.* Retrieved August 15, 2011, from <http://www.educationcounts.govt.nz/indicators/definition/student-engagement-participation/3929>
- Neilson, W. (2000). Disability: Attitudes, history and discourses. In D. Fraser, R. Moltzen & K. Ryba (Eds.), *Learners with special needs in Aotearoa New Zealand.* Palmerston North, NZ: Dunmore Press.
- Parsons, C. (1999). *Education, exclusion and citizenship.* London: Routledge.
- Parsons, C. (2009). *Promoting strategic alternatives to exclusions from school: A development project: Final Report to the Esmée Fairbairn Foundation* (pp. 74). Canterbury: Christ Church University.
- Parsons, C. (2011). *Strategic alternatives to exclusion.* Stoke on Trent: Trentham Books.
- Pomerantz, K. (2007). Containing the emotional needs of adolescent boys 'at risk of exclusion' in mainstream schools. In K. Pomerantz, M. Hughes & D. Thompson (Eds.), *How to reach 'hard to reach' children.* Chichester, UK: John Wiley.
- Skidmore, D. (2004). *Inclusion. The dynamic of school development.* Maidenhead, UK: Open University Press.
- Slee, R. (1995). *The changing theories and practices of discipline.* London: Falmer Press.
- Teachernet. (2010). *Fixed term exclusion* Retrieved 13 January, 2011, from <http://www.teachernet.gov.uk/docbank/index.cfm?id=12255>
- University of Helsinki. (2009). *Centre for Activity Theory and Developmental Work Research.* Retrieved September 8, 2011, from <http://www.edu.helsinki.fi/activity/pages/chatanddwr/chat/>

Enhancing Learner Outcomes Through the Use of Rubrics in the Summative Evaluation Process

Trudy Harris, Dorothy Spiller, Michele Schoenberger-Orgad
and Cheryl Cockburn-Wootten
University of Waikato, Hamilton

Email: scooby@waikato.ac.nz

Abstract

The institutional use of summative evaluations does not generally focus on professional development. Institutions utilise summative evaluations for promotion and other quality requirements. Within this quality environment it is sometimes difficult to get staff to engage with the professional development opportunities offered by the summative evaluations. Furthermore, institutions place a high premium on research especially since the advent of the Performance Based Research Fund (PBRF) exercise. This research focus may also distract academic staff from development around teaching.

In this context, academic developers have to find new ways to improve the link between student feedback and learner outcomes. The literature suggests that interventions around summative evaluations could improve their use for professional development. This paper describes and evaluates the initial stages of a pilot project of such an intervention at the University of Waikato, New Zealand. Specifically, rubrics were introduced to aid interpretation of summative evaluation data and suggest professional development pathways. The initial study, involving a small sample, indicated that there was a positive disposition towards the rubrics and a small number actively engaged with them. The researchers conclude that further iterations of the research are needed to justify the introduction of rubrics, and also argue that such an intervention requires a significant cultural change around teaching to achieve large-scale impact

Introduction

Currently, institutions have a strong focus on summative evaluation to monitor teaching and to identify teaching competence for promotion purposes. The instrument that is used in our context is based around a core of eight questions concerning the paper and the teaching. Students rate teachers on a five point behavioural observation scale. The instrument also provides space for students to make qualitative comments about the teaching and the paper. However, for promotion within the institution, the only information that is required around teaching is the response to one question, overall quality of the paper and the teaching. The newly introduced online promotions process known as the Academic Staff Portfolio (ASP) has accentuated this use of the data, as it is the only form of data collected in relation to teaching. Within this environment academic developers need to devise strategies to enable teachers to maximise the learning potential available from summative evaluations.

The gap between quality and development that is noted here needs to be placed within the context of a number of other environmental factors. One of these is the current preoccupation with the measurement of quality at the governmental and consequently at the institutional level. At the University of Waikato, teaching is governed by a number of strategic documents that set out the University's requirements around teaching. For example, in 2010 the University of Waikato developed the Teaching and Learning Framework, a strategic document that outlined the important areas that staff should engage with to provide students with a quality teaching and learning experience. The framework identifies seven areas that help to provide a quality teaching and learning experience, and offers a rationale and pedagogical underpinning for each area. The framework was then distilled into the Teaching and Learning Plan which contains goals

and key performance indicators (KPIs) for each of these seven areas. One of the KPIs relates to appraisal data and provides a benchmark for all departments/schools/faculties within the institution. The KPIs which the institution has identified provide a starting point for improving the quality of the student learning experience. The challenge for the university is to ensure that the requirement to meet these KPIs does not just become a numerical exercise, but is also complemented by appropriate and relevant developmental support and activity. This problem is exacerbated by the fact that the professional development around teaching is **not** compulsory and there are no minimal requirements around teaching for novices or more experienced academic staff. Furthermore, the Performance Based Research Funding exercise (PBRF), by contrast, now makes it imperative for all academic staff to demonstrate research competence.

The project that will be discussed in this paper is an attempt to encourage academics to make optimal use of student feedback from summative evaluation to improve their teaching and student learning outcomes. Rubrics were introduced around the summative evaluation questions in a pilot study in one department to begin to achieve this goal.

Literature Review: The Connections Between Summative Evaluations and Improving Learner Outcomes

Recent research suggests that summative evaluation data alone will not lead to improvement in teaching quality (Kember, Leung, & Kwan, 2002). Indeed Kember et al.'s work shows that a form of intervention around teaching has to occur before an improvement can be made. An intervention, in this case, could be any form of professional development around teaching. The lack of engagement with summative evaluations may be compounded by perceptions of academic staff. Questionnaire responses in a yet unpublished Ako Aotearoa research study on academics' perceptions of summative evaluation are elucidating in this respect:

Don't even think about it. Too busy to obsess about my teaching. I am a very busy researcher and have a huge teaching and supervision load; I do not have the luxury of time for introspection about teaching.

It is also argued that shortcomings in the summative evaluation process can deter even those who are willing to use the data to inform their teaching development. For example:

The institution only uses the data for quantitative matters. More emphasis needs to come on how individuals can use the appraisal [summative evaluation] system to improve and develop their teaching. Currently, appraisals [summative evaluations] are compulsory, but interpreting them is not compulsory so teachers tend to justify a particular appraisal rather than trying to interpret the results.

Correspondingly, it is suggested that improvements around the summative evaluation process itself could heighten the potential of this instrument for professional development use. In a paper by Spiller and Ferguson (2011), drawing on interviews conducted at the University of Waikato, the qualitative feedback from the respondents showed that there were problems with interpretation of data. For example, one respondent noted that staff should "be taught about interpreting the data and the modifications to make to their courses in response." In support of these findings, work by Smith (2008) suggests that academic staff with little or no pedagogical knowledge had difficulty in interpreting summative evaluation information, and specifically had issues in relating the data to pedagogical areas. Additionally, researchers (Arthur, 2009; Penny & Coe, 2004) found that summative evaluation data themselves often do not suggest pathways for improvement. It is therefore understandable that the absence of forward-looking guidance could deter staff from using summative evaluation feedback as a tool for teaching development.

Drawing on the literature (Smith, 2008) and recognising the perceived limitations of the instrument, teaching development staff at the University of Waikato began to consider ways of enhancing the developmental opportunities that are potentially available in summative evaluation information. Subsequently, a decision was made to embark on a small pilot intervention in collaboration with a department in the Waikato Management School (WMS). The aim of the project was to investigate whether the inclusion of rubrics with the summative evaluation results could enhance the usefulness of summative evaluation as a professional development tool.

Method

In order to investigate whether the inclusion of rubrics in summative evaluation improved the professional development potential of summative evaluations, a pilot study was set up with a selected department in WMS. This particular department was selected as a research partner because of its perceived dedication to teaching. The department has thirteen full-time academic staff, two of whom became research partners on this project. An application for ethical approval was submitted to the WMS Ethics Committee, following the required format. The Committee supported the application and ethical clearance was given, after which we invited all remaining teaching staff to participate in the project. One of the main criteria for participation in the project was that academic staff be teaching in one or both semesters of 2011. Demographics were not an issue in the selection of the participants. Out of the thirteen potential participants two were on study leave over the duration of the project, one was on teaching buy-out, and seven agreed to participate. At this stage it was decided to limit the use of the rubrics to staff before we investigated introducing corresponding rubrics for the students.

The first phase of this project was to align teaching criteria with the expectations formulated in the Teaching and Learning Plan (2010-2012). Drawing on the benchmarks for effective practice identified in the plan, teaching and learning categories were developed for the evaluation questionnaires. For example, around the category related to assessment and resources to support learning, we used the goals and KPIs from Goal 4: Student Assessment of the Teaching and Learning Framework to provide the framework for the related rubric.

Once the main teaching and learning categories had been identified for the questionnaires, it was then possible to start work on the interpretation of the summative evaluation data. After a presentation by Davidson (2010) on the use of rubrics, it was decided to develop a set of rubrics for each of the categories (see Appendix A). The categories are shown in Table 1.

Table 1: Teaching and Learning Categories

Paper appraisal	Teaching Appraisal
Outcomes/Goals	Communication
Assessment and resources to support learning	Facilitation of learning
Intellectual content: Coverage/depth/organisation	Approach/style
Quality of paper	Effectiveness of teaching

Once the rubrics were identified, it was then necessary that attention be paid to the link between the rubrics and possible sources of professional development. A resource document was developed that outlined short-term and longer-term forms of professional development around teaching (see Appendix B).

At the end of Semester A, the questionnaires for this department were integrated into the summative evaluation process and treated as normal summative evaluations. After processing, the appraisal reports were printed for distribution and accompanied by a letter,

the rubrics, and the resources document. This combined package was sent to each participant through the university's internal mail system.

The letter provided personalised guidance to the participants around their paper and teaching and suggested to the participants how they could use the rubrics and the resource sheet to engage in follow-up professional development (see Appendix C). This approach was chosen to give a concrete and contextualised example which would be relevant for each participant.

Advance notice was also given to participants, via the letter, that they would be approached to conduct follow-up interviews to discuss their perceptions of the usefulness of this process. It was decided not to interview the two co-researchers in the department because of their proximity to the design of the project, but their personal reflections were still elicited and recorded separately from the main findings. We felt that our interview data could be influenced by the inclusion of the views of our co-designers.

The interviews were semi-structured around four key questions. It was decided to use prompts to elicit further comment to cater for individual differences. The questions were:

1. How have you used your summative evaluation results for professional development prior to the project? If you have used them please provide specific examples?
2. Have you undertaken or considered further professional development as the result of the project?
3. Describe any differences in your engagement with the summative evaluation results, as a result of the inclusion of the rubrics and resources.
4. What suggestions would you make about improving this tool for professional development?

A thematic analysis of the interview comments would be conducted once the interviews were completed and transcribed. The personal reflections were considered only in relation to the themes arising from the interviews.

A further cycle of this method will be applied at the end of Semester B, incorporating changes and suggestions identified in the findings from Semester A.

Results

The thematic analysis of the data revealed some key themes in relation to the intervention.

Typical Use of Summative Evaluation

Project Participants

In response to this question we found the entire spectrum of views ranging from very detailed usage of comments ("read comments carefully") to lack of interest in summative evaluation ("do not really use appraisal [summative evaluation]"). Within this range, three out of five participants made some use of summative evaluation to inform their teaching practice and course refinements. As a qualifier, it should be noted that the two participants who did not use summative evaluation as a primary tool to inform their teaching both used regular informal feedback and dialogue strategies, and adapted teaching accordingly.

The other participants did not specifically mention their use of other forms of evaluation as this was not the focus of their feedback.

Reflections of a Co-researcher

One of our co-researchers reported “a high level of engagement with qualitative comments.” She also indicated that “appraisal [summative evaluation] feedback has directly motivated extended professional development”

Institutional Uses

Also evident in these responses is reference to institutional use of the summative evaluation by four of the participants. Two of these participants expressed concern because of the consequences that hinge on it, while two others noted that teaching is not rewarded by the institution.

Emotional Responses

Another interesting theme that emerged in relation to current usage was the high level of emotion registered by two participants. Their language indicated their feeling of emotional sensitivity in relation to receiving summative evaluation feedback:

Bit of a fear factor. When I see students going to fetch their assignments and look at their faces, I know how they feel. I think that's how I feel when the brown envelope comes.

Spent an upset night after reading the appraisal [summative evaluation] comments.

Use of Rubrics

Project Participants

There was a generally positive response to the notion of rubrics, although only one of the five academics actively and deliberately used the rubrics to help interpret the feedback from students. For this participant, the rubrics provided a number of valuable functions. The emotional undercurrent, already noted, was also present in this participant's discussion around the use of rubrics. She observed “they acted as a balance or check, calmed my initial response to their comments.” In some instances, the rubrics helped to confirm what she already knew. She also valued the breakdown of results into teaching and learning categories, not only because it made the feedback “less overwhelming” but also because the rubrics enabled her to “highlight the specific areas” that she needed to go back and review. This academic indicated that she would seek out professional development specifically around the areas that the rubrics identified as potentially problematic.

Two academics expressed an appreciation of the potential benefits of the rubrics, and one noted that they were useful for identifying the reasons for lower scores and could sharpen the data. While the other participant found the interpretation “useful”, the emotional theme was evident in her reaction to lower scores, and appeared to block further engagement with the rubrics.

Reflections of a Co-researcher

The generally positive response to the rubrics in the findings was echoed by a co-researcher who indicated that “there has been a difference in my engagement with the results because I found that I have read them far more carefully and referred to the rubric attached.”

Suggestions for Improvements

Project Participants

Some participants provided specific details on improvements to enhance the potential for professional development. For example, one academic said “I’d put a box in relation to where you are in a specific area – and make specific appropriate PD suggestions e.g., helpful reading or attendance at a particular workshop.” Likewise, another participant recognised that the resources needed to be targeted “...and should focus on issues that resonate with them in the classroom.”

For other participants, the view was that the rubrics were limited in their usefulness because of broader issues around summative evaluation and teaching. For one, he said that the formative potential of summative evaluation could only be improved “as part of a larger process in which questions are tailored to different discipline areas.” Two academics again referred to institutional problems around the use of the questionnaires and attitudes to teaching: “...the institution is dysfunctional. Says it uses a TQM methodology, but doesn’t follow through with it.”

Another colleague did not engage with ideas for possible improvements as she felt that teaching is undervalued and that “...appraisal [summative evaluation] has been high-jacked by promotions.”

Reflections of a Co-researcher

Like some of the participants in the research, one of the co-researchers felt the usefulness of the rubrics for professional development was limited by problems with the summative evaluation questions themselves:

I think this is where most of the issues for teachers lie – in the construction of the questions in the appraisals [summative evaluation] and then the students’ choice of answer.

Discussion

Feedback from the research provides some pertinent suggestions on ways of using the summative evaluation student feedback to enhance professional development and learner outcomes. In keeping with Kember et al.’s (2002) view that further intervention around evaluations is needed to make change happen, it was noted that the respondents in this research did not actively use summative evaluations for professional development prior to this initiative. In terms of the use of rubrics, there was a generally positive disposition towards the experiment. However, only one participant in the sample and a co-researcher purposefully employed the rubrics to aid the interpretation of student feedback and apply the findings to future professional development.

Particular suggestions for improvements of the tool as an instrument for professional development related to the creation of more targeted resources to enhance the identified pedagogical shortcomings. The comments around the tool will shape refinements for the next iteration of the project. But a strong message from the results is that while small improvements such as the rubrics may help individuals to enhance their practice, the institutional context around teaching still apparently continues to limit wholehearted and widespread engagement with summative evaluation as a teaching and learning tool. There were also messages about the institution privileging research over teaching, the constraints imposed by the use of summative evaluation for promotion, and serious reservations about the quality of the summative evaluation instrument itself.

Associated with these contextual concerns, this small study has strengthened understanding of the emotional barriers that impede staff from engaging with summative evaluation. Interestingly, this emotional dimension has emerged more vividly in the context of this project than in the analysis of interviews at Waikato (Spiller & Ferguson, 2011).

Conclusion

It needs to be acknowledged that the research was undertaken with a small sample of academic staff in a department that is positive about teaching. The results therefore need to be read with caution. It is also recognised that this is a single intervention and that it needs to be complemented by a raft of other interpretive and developmental tools. Another limitation is that only the teachers and not the students had access to the rubrics. The project initiators will also be considering trialling corresponding rubrics with the students in the future. A further iteration of this project will be run with the pilot group and subsequently introduced to a much wider and more representative selection of academics.

It is important to continue to improve usefulness of the summative evaluation instrument for this goal of professional development, and to make engagement with it as easy and manageable as possible. Furthermore, institutional measures of quality, which include the KPIs of the Teaching and Learning Plan and the use of the ASP, must be directly related to requirements and support for development. A supportive developmental institutional climate may also help to alleviate some of the emotions that academics experience around summative evaluations. In the feedback from participants, it was noticed that while academics were positively disposed towards learning from student feedback there was no evidence of systematic and deliberate reference to student evaluations in order to improve student learning outcomes. It may be speculated that the absence of compulsory requirements around teaching may also contribute to the lack of rigorous attention given to student feedback in this instance. On the basis of this small-scale study it is suggested that interventional strategies such as rubrics may prompt some individual reflection and engagement with student feedback but the link between quality, professional development and student learning must be reflected in all institutional policies and processes for widespread change to happen.

References

- Arthur, L. (2009). From performativity to professionalism: Lecturers' responses to student feedback. *Teaching in Higher Education*, 14(4), 441-454.
- Davidson, E. J. (2010, November). *Actionable evaluation for real-world decision makers*. Paper presented at the 10th Australasian Higher Education Evaluation Forum, Lincoln, NZ.
- Kember, D., Leung, D. Y. P., & Kwan, K. P. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Journal of Assessment & Evaluation in Higher Education*, 27(5), 411-425.
- Penny, A., & Coe, R. (2004). Effectiveness of consultation on student ratings feedback: A meta-analysis. *Review of Educational Research*, 74(2), 215-253.
- Smith, C. (2008). Building effectiveness in teaching through targeted evaluation and response: Connecting evaluation to teaching improvement in higher education. *Assessment and Evaluation in Higher Education*, 33(5), 517-533.
- Spiller, D., & Ferguson, P. (2011, July). *Student evaluation: Do lecturers value them and use them to engage with student learning needs?* Paper presented at the Education in a Changing Environment 6th International Conference: Creativity & Engagement in Higher Education, Salford, UK.

Appendix A

Paper and Teaching Rubrics – The Paper

Pedagogical Area	Questions	Excellent 1 – 1.5	Very Good 1.6 – 2.0	Satisfactory 2.1 – 3.0	Unsatisfactory 3.1 – 4.0	Unacceptable 4.1 – 5.0
Outcomes/ goals	1. I received accurate information about what the paper would involve 2. The paper's objectives (learning outcomes) were stated clearly	Learning outcomes are explicitly aligned with content, teaching & learning strategies and assessment.	Alignment between learning outcomes, content, teaching & learning strategies and assessment is evident, but minor omissions.	Some alignment between learning outcomes, content, teaching & learning strategies and assessment.	Poor alignment between learning outcomes, content, teaching & learning strategies and assessment.	No explicit alignment between learning outcomes, content, teaching & learning strategies and assessment.
Assessment and resources to support learning	3. The assessment tasks were a fair test of my learning in this paper 4. I was able to access required learning resources when I needed them	Assessment facilitates relevant learning; Sets out well-defined criteria for successful completion; Provide access to excellent resources.	Assessment is linked with learning and has clear criteria for successful completion; Good resources.	Assessment has reasonable links with learning and rudimentary criteria for successful completion; Adequate resources.	Assessment has marginal links with learning and with ill-defined criteria for successful completion; Some resources.	Assessment has no link with learning and no criteria defined for successful completion; minimal resources.
Intellectual content	5. The paper was well organised and ran smoothly 6. I was given enough time to learn the things I had to learn 7. The paper was intellectually stimulating	Well crafted structure and design of the paper. Organisation of the paper was explicit. Excellent balance between content and learning processes.	Coherent structure, organisation and design to the paper. Good balance between content and learning processes.	Basic structure, organisation and design to the paper. Some attention is paid to the relationship between content and learning processes.	Rudimentary structure, organisation and design to the paper. Scant attention is paid to the relationship between content and learning processes.	No structure, organisation or design to the paper. No attention is paid to the relationship between content and learning processes.
Quality of the paper	8. Overall, I was satisfied with the quality of this paper					

Appendix A

Paper and Teaching Rubrics – The Teaching

Pedagogical Area	Questions	Excellent 1 – 1.5	Very Good 1.6 – 2.0	Satisfactory 2.1 – 3.0	Unsatisfactory 3.1 – 4.0	Unacceptable 4.1 – 5.0
Communication	<ul style="list-style-type: none"> • This teacher was good at explaining things • This teacher made it clear what they expected of me 7. This teacher was approachable when advice or help was required	Established clear expectations and builds dialogue throughout. Uses accessible language and a learner responsive pace.	Expectations are communicated. Develops dialogue between teachers and students. Accessible language and learner responsive pace.	Rudimentary expectations are established. Beginning of dialogue between teachers and students. Appropriate language and pace.	Expectations are unclear. No dialogue between teachers and students. Language is still difficult and pace imperfect.	Expectations are not communicated. No dialogue between teachers and students. Language is inaccessible and at an erratic pace.
Facilitation of learning	4. This teacher gave me helpful feedback on how I was going 5. This teacher encouraged me to get actively involved in learning tasks	Always responsive to learners and actively promotes the development of intellectual independence. Facilitates environment for reflection and practice.	Responsive to learners and develops intellectual independence. Environment provides opportunities for reflection and practice.	Aware of learners' needs and begins to encourage intellectual independence. Environment provides limited opportunities for reflection and practice.	Little awareness of learner needs and does not promote intellectual independence. Environment neglects opportunities for reflection and practice.	Ignores learner needs and stifles development of intellectual independence. Environment discourages reflection and practice.
Approach/style	1. This teacher was enthusiastic about what they were teaching 6. This teacher made the subject interesting	Excites curiosity and actively seeks dialogue and interaction. Teacher uses a range of pertinent strategies and examples.	Invites curiosity and encourages dialogue and interaction. Uses a range of strategies and appropriate examples.	Primarily transmission of information & begins to interact with students. Some variation in strategies and occasional use of examples.	Transmission of information & minimal interaction. Very limited strategies and use of examples	Monotonous transmission of information and no interaction with students
Effectiveness of teaching	8. Overall, this teacher was effective					

Appendix B

Paper and Teaching Resources – The Paper

Pedagogical Area	Questions	Short Term Professional Development		Long Term Professional Development	General Resources
		TDU Resources	University Resources		
Outcomes/ goals	<ol style="list-style-type: none"> 1. I received accurate information about what the paper would involve 2. The paper's objectives (learning outcomes) were stated clearly 	TDU Handbooks <ul style="list-style-type: none"> • Tertiary teaching: Exploring our Beliefs • Introduction to Course Design • Becoming a Reflective Practitioner • Maximising Learning in Large Groups 	Teaching and Learning Framework https://i.waikato.ac.nz/portal/server.pt/directory/university_plans/10622?DirMode=1		AKO Aotearoa http://ako.aotearoa.ac.nz/
Assessment and resources to support learning	<ol style="list-style-type: none"> 3. The assessment tasks were a fair test of my learning in this paper 4. I was able to access required learning resources when I needed them 	<ul style="list-style-type: none"> • Principles of Assessment • Assessment tasks to promote learning • Assessment: Feedback to Promote Student Learning • Assessment: Setting and Marking Assessment Tasks 	Teaching and Learning Plan http://www.waikato.ac.nz/about/corporate/tlmgplan.shtml	PGCert in Tertiary Teaching http://www.waikato.ac.nz/tdu/PGCERT.shtml	The Higher Education Academy http://www.heacademy.ac.uk/
Intellectual content	<ol style="list-style-type: none"> 5. The paper was well organised and ran smoothly 6. I was given enough time to learn the things I had to learn 7. The paper was intellectually stimulating 	<ul style="list-style-type: none"> • Assessment: Setting and Marking Assessment Tasks • Tutoring: A Guide to Getting Started • Facilitating Laboratory and Practical Classes 	Paper Outline Policy http://www.waikato.ac.nz/official-info/index/docs/paper-outline-policy	Mentoring http://www.waikato.ac.nz/hrm/mentor/	Phil Race website: http://phil-race.co.uk/ (Phil was a visiting scholar at UoW in 2010)
Quality of the paper	<ol style="list-style-type: none"> 8. Overall, I was satisfied with the quality of this paper 	Teaching Network: http://www.waikato.ac.nz/tdu/tchgntwk.shtml TDU workshops: http://www.waikato.ac.nz/tdu/staffworkshops.shtml TDU individual support TDU Library	Policy on the Evaluation of Teaching and Papers http://www.waikato.ac.nz/official-info/index/docs/evaluation-of-teaching-and-papers-policy		Stephen Brookfield website: http://www.stephenbrookfield.com/Dr_Stephen_D_Brookfield/Home.html
		Further information and resources can be found at the TDU website: www.waikato.ac.nz/tdu Or contact Preetha Pratapsingh at: tduadmin@waikato.ac.nz ; telephone 07 838 4839			

Pedagogical Area	Questions	Short term Professional Development		Long term Professional Development	General Resources
		TDU Resources	University Resources		
Communication	2. This teacher was good at explaining things 3. This teacher made it clear what they expected of me 7. This teacher was approachable when advice or help was required	TDU Handbooks <ul style="list-style-type: none"> Tertiary teaching: Exploring our beliefs Becoming a Reflective Practitioner Teaching Strategies to Promote the Development of Students' Learning Skills Teaching International Students Maximising Learning in Large Groups Assessment: Feedback to promote student learning Assessment: Setting and Marking Assessment Tasks 	Teaching and Learning Framework https://i.waikato.ac.nz/portal/server.pt/directory/university_plans/10622?DirMode=1	PGCert in Tertiary Teaching http://www.waikato.ac.nz/tdu/PGCERT.shtml	AKO Aotearoa http://ako.aotearoa.ac.nz/ The Higher Education Academy http://www.heacademy.ac.uk/
Facilitation of learning	<ul style="list-style-type: none"> This teacher gave me helpful feedback on how I was going This teacher encouraged me to get actively involved in learning tasks 		Teaching and Learning Plan http://www.waikato.ac.nz/about/corporate/tlmgplan.shtml	Mentoring http://www.waikato.ac.nz/hrm/mentor/	Phil Race website: http://phil-race.co.uk/ (Phil was a visiting academic at UoW in 2010)
Approach/ style	1. This teacher was enthusiastic about what they were teaching <ul style="list-style-type: none"> This teacher made the subject interesting 	Teaching Network: http://www.waikato.ac.nz/tdu/tchgntwk.shtml	Paper Outline Policy http://www.waikato.ac.nz/official-info/index/docs/paper-outline-policy	Peer Observation http://www.waikato.ac.nz/tdu/observations.shtml	Stephen Brookfield website: http://www.stephenbrookfield.com/Dr_Stephen_D_Brookfield/Home.html
Effectiveness of teaching	8. Overall, this teacher was effective	TDU workshops: http://www.waikato.ac.nz/tdu/staffworkshops.shtml	Policy on the Evaluation of Teaching and Papers http://www.waikato.ac.nz/official-info/index/docs/evaluation-of-teaching-and-papers-policy		
		TDU individual support TDU Library			
Further information and resources can be found at the TDU website: www.waikato.ac.nz/tdu Or contact Preetha Pratapsingh at: tduadmin@waikato.ac.nz ; telephone 07 838 4839					

Appendix C: Letter to Academic Staff

Dear [Academic staff member],

You are receiving this letter as part of the research project looking at interpretation of appraisal data and linkages to professional development around teaching. We would like to take this opportunity to thank you for taking part in this research project, and to explain about the next step in the project.

This part of the project is to provide you with information which will enable you to make informed choices around professional development activities. Enclosed with this letter are a number of documents, including your appraisal results and rubrics for the interpretation of the appraisal. The data for the appraisal questions has been aggregated under the relevant pedagogical areas. For each of these areas rubrics have been developed, in line with the University's Teaching and Learning Framework. These rubrics outline what excellent to poor teaching activity would look like for each of the pedagogical areas. This makes it easier to align professional development offering and opportunities.

So for example, for your papers the scores are listed below:

Paper code	Pedagogical Areas							
	Outcomes/goals	Assessment and resources to support learning	Intellectual content	Quality of the paper	Communication	Facilitation of learning	Approach/style	Effectiveness of teaching
TDU100/200-11A (HAM)	1.8	2.0	1.9	2.1	1.7	1.9	1.8	1.8
TDU585-11A (HAM)	1.5	1.6	1.3	1.3	1.3	1.1	1.0	1.1

Paper appraisal

Teaching appraisal

The rubrics document on the next page outlines for you what each of the scores means in terms of teaching practice. For example the scores for TDU585 clearly indicates that, for Outcomes/goals and Intellectual content your teaching practice is excellent, however for Assessment and resources to support learning there is potential for further work.

We would really appreciate hearing your views about the process and the use of rubrics and would like to contact you at the end of July to arrange an interview.

If you have any questions please contact me or your departmental representatives on this project Cheryl Cockburn-Wootten and Michele Schoenberger-Orgad.

Kind regards

Design Experiments for Assessment of “21st Century” Learning Outcomes

Rosemary Hipkins
New Zealand Council for Educational Research

Email: rose.hipkins@nzcer.org.nz

Abstract

Debates about educational outcomes that equip learners for life in the complex and rapidly changing employment and life contexts of the 21st century focus on ideas such as learning-to-learn and the development of key competencies. The participatory nature of learning is emphasised but this begs some interesting questions concerning relationships between knowledge and action. What are students expected to *do* with their learning and what are the implications for assessment? Assessment Resource Bank (ARB) items, designed with assessment for learning in mind, are typically developed via small design experiments that reframe traditional assessment tasks with the New Zealand Curriculum key competencies in mind. Item analysis seeks to unravel the thinking behind students' responses and hence to suggest ways teachers might use the tasks to inform further opportunities to develop and strengthen key competencies in subject-specific contexts. This paper presents a simple ARB science classification task completed by 340 year 6 and 8 students from a range of schools. The results point to participatory opportunities for students to engage in criterion-based classification inquiries, justifying their conclusions with reference to the evidence available, and challenging each other's claims in the same manner that scientists employ in their own knowledge-building activities.

Introduction: The Curriculum Challenge

In common with many other nations, New Zealand is wrestling with questions of what it means to educate students for the rapidly changing economic, environmental and social conditions that characterise life in the 21st century (Bolstad & Gilbert, 2008; Gilbert, 2005). The most recent New Zealand Curriculum (NZC) is a future-focused *framework* curriculum whose purpose is to provide a sense of national direction for local decision-making (Ministry of Education, 2007). Each school has to work out how best to build up a detailed local curriculum based on the national framework, with the identified learning needs of its own student community demonstrably addressed (Hipkins, Cowie, Boyd, Keown, & McGee, 2011).

A vision statement and a set of principles guide the reading and interpretation of the whole. The vision is for students to become “confident, connected, actively involved lifelong learners” (Ministry of Education, 2007, p. 8). The vision and principles are given life when schools design learning programmes that weave more traditional content with specified values and key competencies. Eight broad sets of values, identified and shaped via a national consultation exercise, are expected to be encouraged, modelled and explored. The five NZC key competencies were adapted from a set of four developed by the OECD's DeSeCo project (Rutherford, 2005). DeSeCo defined as “key” those competencies students need to develop during their schooling in order to maximise their chances of living meaningfully in, and contributing to, well-functioning societies, both during and well beyond their school years (OECD, 2005). Learners draw on a wide range of competencies, but those labelled as key are seen to be universal rather than situation specific (Rychen & Salganik, 2003). The implication is that these competencies are transferrable across contexts and continue to develop across the lifespan.

Congruent with the OECD intent, NZC defines the key competencies as “capabilities for living and lifelong learning” and notes that their development is a *means to other ends* as well as an end in itself (Ministry of Education, 2007, p. 12). The use of the term

capabilities in the NZC definition statement, especially in conjunction with the notion of lifelong learning, signals a focus on what students can do with their learning – what they are capable of *doing and being* now, as well as looking forward to their future capabilities (Walker, 2008). Students are expected to bring appropriate combinations of knowledge, skills, attitudes and values together so they can take action in contexts that strengthen and stretch their current abilities. Underscoring this intent, one of the five NZC key competencies is called “participating and contributing.”

A “capabilities” framing of key competencies (as opposed to more traditional skills-based interpretation) reflects the participatory ethos that pervades arguments for a so-called “21st century” curriculum. In this usage participation is seen as an alternative that complements and enriches “acquisition” as a metaphor for learning (Sfard, 1998). Thus the use of the term participation does not necessarily entail working directly with others during learning (although this will often be the case), but rather emphasises what students can *do* with what they learn. There should be a purpose beyond simply accumulating knowledge for its own sake (for an extended discussion of the nature of participating and contributing as a key competency see Bolstad, Roberts, Boyd, & Hipkins, 2009).

Schools face a considerable challenge as they explore ways of integrating key competencies with the more traditional “achievement objectives” of the eight NZC learning areas (Cowie, Hipkins, Keown, & Boyd, 2011; Hipkins & Boyd, 2011). Science teachers have the additional challenge of coming to grips with a new Nature of Science (NOS) strand, which is “overarching” and hence intended to be integrated with the four “contextual” strands that broadly encompass more traditional disciplinary divisions within science (i.e., biology, chemistry, physics, and an earth science/astronomy mix) (Bull, Joyce, Spiller, & Hipkins, 2010). All teachers need access to usable exemplars that show them the difference that key competencies could make to more traditional learning activities. In science they also need to: develop their understandings of the intent of the NOS strand; work out how the changes signalled by this strand might relate to those being signalled by the key competencies; and bring both of those sets of professional learning together to revise and update or create completely new learning experiences for students. This is both conceptually and practically demanding and the onus is on researchers and teacher educators to do anything they can to help.

Investigating in science is one of the sub-strands of the NOS strand in the NZC. By definition, investigation is an active and participatory process that requires learners to be engaged and purposeful, whether they are exploring alone or with others. Investigation has been an important component of science learning ever since its inception as a formal part of the school curriculum (DeBoer, 1991) but its nature as a learning activity has also been the subject of ongoing debate (see for example Millar, 1998). This paper does not have the space to canvass those ongoing arguments; the point being made is that it is not self-evident how teachers should structure and support student investigations so that NOS insights are one outcome of learning (Ford & Forman, 2006). In this contested terrain, NZC gives no explicit guidance as to how primary school children’s investigative experiences might more deliberately scaffold the development of key competencies, let alone begin building a sense of what science is like as a knowledge-building discipline (i.e., inform NOS understandings). The relevant NOS objectives on the fold-out sheets in NZC (Ministry of Education, 2007) are:

Extend their experiences and personal explanations of the natural world through exploration, play, asking questions and discussing simple models. (Levels 1 and 2)

Build on prior experiences, working together to share and examine their own and others’ knowledge. Ask questions, find evidence, explore simple models, and carry out appropriate investigations to develop simple explanations. (Levels 3 and 4)

The focus is on exploring and sharing ideas by talking and investigating together. But how might students do this in an age-appropriate manner that conveys an authentic sense of

“working like a scientist?” What would evidence of their learning success look like? What new insights might they take away from their learning that could be useful for later thinking and action in the world? (Ford & Forman, 2006). This is the curriculum and assessment design dilemma that this paper addresses.

The Design Opportunity: Leveraging Existing Work Streams

Assessment Resources for Classroom Teachers (ARCT) is a contract between the New Zealand Council for Educational Research (NZCER) and the New Zealand Ministry of Education that includes the development of assessment resources for formative purposes and making these available to New Zealand teachers through a dedicated website, the Assessment Resource Banks (ARB). The tasks are designed for teachers to use as part of their everyday classroom activities. There are many different definitions for formative assessment, but the purpose is common to all – to make pedagogical decisions that improve student learning (Newton, 2007). The use of assessment for such purposes, regardless of the nature of the task, poses particular challenges for teachers. Three main actions are required – eliciting student responses, interpreting what these responses might mean in terms of next learning steps, and acting on these insights (Cowie & Bell, 1999). Writing assessment tasks for formative purposes is challenging because assessment cannot be regarded as formative *until* it results in action by the teacher and/or the student (Black, Harrison, Lee, Marshall, & Wiliam, 2002; Newton, 2007). While knowing whether students can provide the “right” answer is useful information, understanding *how* the student got to their answer provides the platform for deciding what to do next. The tasks, therefore, need to be designed to reflect this requirement. Their recent design experiments have led the ARCT team to believe that these challenges are compounded when the focus is on *using* rather than having knowledge.

Because ARB resources are intended to support teachers to elicit, interpret and respond to students’ learning challenges, each task is accompanied by educative material to assist teachers and students with analysing student responses and making learning and teaching decisions about next steps. Pen-and-paper ARB resources are trialled nationally as sets of three or four items, often related to each other in some way. Around 200 students from a sample of New Zealand schools complete each task booklet and at the end they are given the opportunity to make a personal response about enjoyment and difficulty levels of items. The NZCER science education research team work together to analyse response patterns, first reading and discussing a sample of completed scripts, then building and applying a coding schedule based on the researchers’ emerging understanding of the patterns of responses. Wider research projects in specific areas of interest related to formative assessment complement the more formal tasks from time to time.

All recently created science ARB items have experimented with the integration of a foregrounded key competency with some aspect of curriculum content. Each one is a modest example of a research-based *innovation*. Such innovation seeks first to understand the nature of the problem to be addressed, then to design innovations to address the issues as they are now understood. Important questions asked of research-based innovation include: What is this idea good for? What does it do and fail to do? Does it have a future? How could it be improved? Knowledge creation then “depends on moving back and forth between belief and design questions in ways that maintain progress in idea improvement” (Bereiter & Scardamalia, 2008, p. 79). The idea being explored for the assessment item that is the focus of this paper was biological classification and its potential relationship to NOS insights and key competency development.

The Design Experiment: Adapting a Traditional Assessment Task

Learning about classification is a common focus of primary school science and there are a number of items in the science ARB collection that assess students' current *knowledge* of classification categories. The overarching NZC achievement aim for the investigating in science NOS sub-strand names "classifying and identifying" as a relevant type of investigation. We were interested in adapting some of the existing tasks to reflect more clearly the participatory intent of NZC. The design question here is: what can children *do* with their current knowledge and understandings of classification? The challenge that NOS adds is to ensure that the investigative experience of classifying also conveys a sense of what it is like to classify in a "scientific" manner.

A recently described model addresses the participatory NOS challenge by identifying two separate but interrelated roles that children need to experience as they investigate scientifically (Ford & Forman, 2006). Science aims to create new knowledge about the natural and physical world, and to do so in ways that are convincing and carry authority. That is the very essence of its nature as a *discipline*. Scientists need to convince their peers in the first instance, and then others, that their new ideas should be adopted to either improve or replace what has gone before. They cannot do this just by the force of their personality or by cheating in some way. Some might try but they ultimately get found out because there are "rules of the game" that they need to follow if they hope to have their work elevated to the status of new knowledge. These rules will obviously be different in their specifics depending on the field of science involved (methods are different for a start) but Ford and Forman suggest that they can be reduced to the two main areas of activity: constructor of claims and critiquer of claims. You cannot claim to *be* a scientist unless you can do both successfully *and in combination*.

Scientists work to make nature "speak" in ways that they can then document. In the role of a *constructor* of claims they need to justify any claims they make with reference to evidence. Simplifying situations to get rid of distracting complexities, manipulating the aspect of interest and then measuring and describing what happens are at the heart of scientists' work in the constructor role. In this role scientists must also show how their ideas relate to what is already known in their field. Whether they accept or reject existing science theories of relevance to their work, they will not be taken seriously if they do not relate these theories to their argument. In the role of a critiquer of the claims of others, scientists need to challenge their peers to ensure they do the same. New science claims won't survive long unless they are convincing to a scientist's peers. Testing the accounts of others for any flaws is at the heart of the critiquer role.

This model suggests clear opportunities for rich conversations, as signalled in the achievement objectives cited above, but what sort of evidence-based claims might children be able to make in relation to classification of the living world? Unless there is something interesting and complex to discuss (Gallas, 1995), there will be no opportunity to experience the constructor and critiquer roles, or the manner in which they interrelate. One possibility here is to introduce children to the manner in which scientists use agreed criteria when making decisions about how to classify a specific organism. A challenging but potentially interesting feature of such investigations is that both confirming and disconfirming evidence must be accounted for. Furthermore, the evidence available will be incomplete if one or more of the criteria cannot be either confirmed or disconfirmed by the observations that can be made. There are decisions to make and justifications to present when children classify by checking their careful observations against a list of established criteria. One possible way to scaffold investigative conversations about classification is to give children a clear signal that the exercise is not simply about producing a right answer. In the case of the task described next this signal entailed adding an "I can't tell" option to an otherwise relatively simple classification task.

For this assessment task, students were provided with a list of features characteristic of the reptile family of animals. They have in common:

- a backbone
- cold-blood (temperature adjusts with temperature of surroundings)
- breathe with lungs
- a dry skin covered with scales
- either 4 legs or none, and
- ear holes but no ears.

Many also have a hard shell and most lay eggs which have a strong soft shell, although a few reptiles have live babies. As well as a simply worded list of these features students were given photographs of seven animals and told the task was about “working like a scientist to decide whether or not an animal is a reptile.” They were asked to *observe* closely and *use* the features on the list to *decide* which of the animals were reptiles, and which were not. For each animal, they circled their response (yes, no, I can’t tell). In a box alongside they were also asked to give reasons for their answer, so that the researchers could gain at least a little insight into their decision-making.

The appropriate time to use “I can’t tell” is when the evidence presented is insufficient to make a determination. However, as we shall see shortly, actually using it in this manner requires students to understand that the purpose of the task is to use and apply evidence, not simply to arrive at the “right answer.” If “I can’t tell” is used as intended there are *dispositional* components to this task. As has been noted in relation to the development of key competencies in general, students need to be “ready, willing and able” to do what is required (Carr, 2006). In this case they need to suspend making a judgment until after they have observed and decided in the manner cued by the question. The task also potentially highlights the role played by disconfirming evidence. An animal that has most but not all features of reptiles will not be a reptile. Thus every item on the provided list must be considered and the task directs attention to differences, as well as similarities. This consideration of all the potential evidence is not completely straightforward in terms of a photograph-based *observation* task because some features cannot be determined without recourse to prior knowledge or additional research.

When considering the results reported next, it is important to keep in mind the limitations of the trial conditions (which are currently limitations of the research contract opportunity being leveraged). On the ARB website the photographs are in full colour; if teachers use an interactive whiteboard or other data display instead of printed copies of the item this is how students will see them. However, for the trial, which was pencil and paper based, students worked with black and white copies. Furthermore, they worked individually rather than in discussion, so their decision-making opportunities were private rather than collaborative. Because the ARB item trials are intended to give an insight into students’ likely current thinking and decision-making (as opposed to their potential capabilities when given access to different sorts of learning opportunities), these limitations need not transfer to teachers’ classroom-based use of the item as a formative assessment resource, or its cooption as a learning activity. On the contrary, the teacher notes that accompany the item use the outcome of the trial to emphasise how teachers might use the item in ways that encourage discussion and debate and hence the development of aspects of participation as a key competency and investigation as a NOS-related dimension of science learning.

Learning from Students’ Responses

The trial was completed by 340 students: 169 were in year 6 and 171 in year 8. Students came from 16 different schools and their own teachers introduced and administered the task booklet. Responses to the first part of the task were captured as basic frequency data (yes, no, I can’t tell, missing). Each reason given in the second part of the task was coded as one of the following: feature that can be observed (A); correct feature but cannot be observed

(B); very general correct statement e.g., “it’s an insect” (C); incorrect reason (D); no reasons given (missing M). These data were captured as combinations of responses. For example, AAA would mean a student gave three observable reasons; AABB would mean they gave two observable and two non-observable reasons; combinations of only B responses, or B, C and D responses would mean they likely drew only on their general knowledge.

The pattern of responses to the first part of the task (yes/no/I can’t tell) is shown in Table 1. Alongside this is our overall judgement of how easy it was for students to arrive at the “right” classification (80-100% very easy, 60-80% easy, 40-60% moderate, 20-40% difficult, 0-20% very difficult). Note that the table has been ordered by the frequency of use of the “I can’t tell” option (the shaded row), not the order in which the animals appeared in the task.

Table 1: Students’ Choice of Classification Option

Animal depicted	Year 6 % (n = 169)			Year 8 % (n = 171)			Is it a reptile? Ease of arriving at “right” answer
	Yes	No	Can’t tell	Yes	No	Can’t tell	
Snake	88	9	2	91	6	2	Very easy
Lizard	91	5	3	93	4	4	Very easy
Bird	4	90	4	4	87	8	Very easy
Turtle	81	9	9	84	6	8	Very easy
Fish	27	65	8	28	58	12	Easy (yr 6) Medium (Yr 8)
Frog	53	37	10	60	28	11	Difficult
Weta	9	75	14	8	79	12	Easy

Rows per year level may not add to 100 because of rounding and missing data

We found no overall difference between the year 6 and year 8 students’ response patterns, with the interesting exception that older students seemingly found it somewhat harder to classify fish as “not reptile”. There was a clear relationship between students’ use of “I can’t tell” – limited as that was – and the ease with which they made a correct choice of category. The easier the animal was to classify, the less likely students were to say they couldn’t tell. The weta was the only animal that did not fit this pattern but it may well be the exception that proves the rule. The most difficult animal for the students to classify was the frog. There were indications that those who chose the right response (not a reptile) did so by drawing on existing knowledge rather than using the list provided:

From what I have learnt at school I’ve learnt frogs are amphibians.

I just know it’s an amphibian.

There were occasional indications that students did use the “I can’t tell” category as intended when classifying the frog (“I can’t tell if it has a dry skin”) but on the whole “I can’t tell” appeared to mean “I don’t know.”

The weta is an interesting example of a small animal that is likely to be familiar to many New Zealand children, but not all of them. Wetas often crawl out of rotten timber when disturbed so it is common for children to see them if they live at or near ground level, especially in suburban or rural settings. New Zealand has traditionally been more suburban than urban, but increasingly our school population includes refugee children and those who live in high-rise apartments where they may never have seen wetas. Thus it seems likely that its classification was easy for students who are familiar with them but perhaps not for those who have never seen them before. Again, some students’ reasons for choosing “not a reptile” support the assertion that they were drawing on prior knowledge, including disconfirming evidence that is clearly visible in the photo:

I *know* a weta is an insect. It has 6 legs, no scales, no hard shell, doesn't have a backbone (emphasis added).

A weta does of course have an exoskeleton, so not quite all of this student's prior knowledge is correct. The main point though is that what the student *already knows* seems to be a key affordance they bring to their learning, doubtless in combination with seeing the purpose of the task as arriving at the right answer. For most of our students this was likely not to be a classification investigation in the sense that it would be necessary to observe closely and make a determination based on presence or absence of features on the provided list. Instead reaching the correct classification seemed to be a knowledge recall task and mostly an easy one at that.

What of the other dimension of the task – citing evidence to support the choice of response? The next two tables show the results of an analysis of the combinations of evidence the students cited. Table 2 shows year 6 students' responses and Table 3 shows those of the year 8 students. From left to right, responses are ordered by the extent to which students drew *only* on their own knowledge (the shaded row of each table), seemingly ignoring the intent of the task when they did so.

Table 2: Year 6 Students' Use of Evidence (n = 169)

Combinations of evidence cited by individuals	Fish	Frog	Bird	Weta	Snake	Lizard	Turtle
% own knowledge	43	36	34	29	27	24	10
% observable plus own knowledge	25	30	25	23	38	34	36
% observable only	16	15	23	22	26	27	34
% missing or other	15	20	18	26	9	15	20

Numbers may not add to 100 because of rounding

Table 3: Year 8 Students' Use of Evidence (n = 171)

Combinations of evidence cited by individuals	Fish	Frog	Bird	Weta	Snake	Lizard	Turtle
% own knowledge only	50	38	31	27	19	19	11
% observable plus own knowledge	23	32	32	26	53	56	48
% only observable	7	15	26	28	21	18	29
% missing or other	20	15	12	18	6	7	12

Numbers may not add to 100 because of rounding

Notice that this ordering is the same at both year levels. All the vertebrate non-reptiles appear towards the left (students drew more on own knowledge) and the reptiles all appear to the right (students cited more of the observable features) with the weta in the middle. Notice too the overall lower frequency of non-responses (missing data) for the reptiles. It appears that students were more likely to resort to evidence from their own knowledge base when they encountered animals that shared some but not all of the features of reptiles – that is, they needed to weigh up the implications of finding *both* confirming and disconfirming evidence. Alternatively, we could say it was simply easier to choose observable evidence for the reptiles since they matched the provided list on all points. However, as already noted, not all of the listed features were actually observable.

Could it be that more year 8 students *did* attempt to engage with the task as intended, notwithstanding the small overall differences in the results? The non-reptilian vertebrates are precisely the animals that the year 8 students appeared to find somewhat more difficult to classify correctly (Table 1). Could it be that some of them were in fact struggling to resolve a tension between completing the observation task signalled and getting the right answer by drawing on their general knowledge? The only way to know for sure would be to discuss their responses with the students themselves – which is what we hope teachers will do when they use this task as a shared investigation that provides rich opportunities for students to play roles as both constructors and critiquers of evidence-based claims.

The next table shows the results of cross-tabulating the first and second parts of the students' responses to the weta photograph. We did this for all seven examples and chose this one for inclusion since more students chose "I can't tell" for this animal. (Note that the table shows actual numbers not percentages.) Confirming the inference that "I can't tell" actually meant "I don't know" for many students, we can see here that those who chose "I can't tell" were more likely to have missing data for the evidence part of the question. For example, 19 of the 23 year 6 students and 15 of the 21 year 8 students who chose "I can't tell" for the weta gave no evidence to the second part of the question.

Table 4: Analysis of Use of "I can't tell" Response for Weta

Evidence cited	Year 6 classification (n = 169)				Year 8 classification (n = 171)			
	Yes	No	Can't tell	missing	Yes	No	Can't tell	missing
Observation only	2	35	-	-	2	45	1	-
Observable plus own knowledge	1	37	1	-	4	39	2	-
Own knowledge only	10	34	3	2	4	40	3	1
No evidence cited	2	20	19	3	3	11	15	1

Overall then, it seems that despite the information we provided and asked students to use, they interpreted this task as being about arriving at the correct classification by any means they could. Many did successfully draw on their existing knowledge, and some did so in combination with the provided list. While drawing on one's prior learning is obviously to be encouraged, these responses suggest that this task, used traditionally, would be unlikely to contribute to competency development by providing an experience of constructing and critiquing knowledge claims in the manner described above. There is a little evidence to suggest that opportunities would be taken up to explore the role that evidence, especially disconfirming or indeterminate evidence, might play in expanding the scope of the task to include competency dimensions, if students work alone. However, if they take part in investigative conversations, as suggested by the NOS sub-strand of the NZC, the aspects that students find hard if they don't "just know" could be easily transformed, with the teacher's support, into opportunities for developing new skills and insights about how classification is enacted in a scientific manner.

From Assessment Design Experiment to Pedagogical Change

It might seem that in designing an assessment task to explore a new pedagogical possibility we have put the cart before the horse. There are precedents and good reasons for doing so. One challenge of systematically redesigning assessment tasks to reflect 21st century learning outcomes better is that it is not necessarily clear in advance what the assessment target should be. Compared with more traditionally assessed knowledge acquisition outcomes, the question of what students can do with their knowledge and skills requires the gathering of somewhat different sorts of evidence (Ford & Forman, 2006), but

first there is a need to determine what might be possible and plausible. Responding to this challenge, Scalise and Wilson (2011) cite earlier design work to advocate for the development of *evidence-centred* assessment architecture. This involves four sequential stages:

- select and develop tasks based on construct;
- present tasks to learners;
- learners generate evidence with respect to constructs;
- evidence used to make inferences about construct of interest.

These are the steps we follow in our ARB design experiments. Ideally, we would have been able to design a group assessment and observed students' investigations in action. However, the outcomes of the pencil and paper opportunity afforded by our ARB contractual commitments can now inform future investigative learning activities in the classroom, with associated opportunities to generate different sorts of assessment evidence. Students might provide evidence of their ability to *construct* evidence-based claims by justifying their decisions with reference to the criteria provided. At the same time, the possibilities of finding disconfirming evidence, or of acknowledging that evidence for one criterion is not available, provide good scope to critique the claims of others (or to anticipate critique of one's own claims). However, the evidence that our design experiment generated clearly suggests that teachers will need to scaffold these possibilities if students are *not* to experience classification tasks as being about finding the right answer.

It has been suggested that the real difference between 20th and 21st century learning resides in the development of a "thinking curriculum" for *all* students, not just the most able as has traditionally been the case (Resnick, 2010, p. 186). Lauren Resnick explains that a thinking curriculum should employ guided classroom discussion of core disciplinary ideas. In this view, learning needs to be infused with an intellectual quality that matters for more than just the task at hand and the teacher needs to bring deep disciplinary knowledge to bear on the emergent learning. If they are to expand their practice repertoires in this way, primary teachers do need examples that show what such teaching and learning might look like, and how disciplinary expertise in the sciences may be brought to bear on the learning in question. This article has suggested that making changes to achieve disciplinary depth may sometimes be as simple as opening up traditional tasks by using a strategy such as "I can't tell" as an opening gambit for a discussion of nature and enactment of classification, in place of the more traditional focus on its products. Precisely because classification is such a common topic, it may be that teachers will feel more confident to experiment with the sorts of changes envisaged than they might be in less familiar science contexts.

References

- Bereiter, C., & Scardamalia, M. (2008). Toward research-based innovation. In Centre for Educational Research and Innovation (Ed.), *Innovating to learn, learning to innovate*. Paris: Organisation for Economic Co-operation and Development.
- Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2002). *Working inside the black box*. London: nferNelson.
- Bolstad, R., & Gilbert, J. (2008). *Disciplining and drafting, or 21st century learning? Rethinking the New Zealand senior secondary curriculum for the future*. Wellington: NZCER Press.
- Bolstad, R., Roberts, J., Boyd, S., & Hipkins, R. (2009). *Kick Starts; Key competencies: Exploring the potential of participating and contributing*. Wellington: NZCER Press.
- Bull, A., Joyce, C., Spiller, L., & Hipkins, R. (2010). *Kick starting the nature of science*. Wellington: NZCER Press.
- Carr, M. (2006). *Dimensions of strength for key competencies*. Retrieved February 2008, from http://nzcurriculum.tki.org.nz/curriculum_project_archives/references

- Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education: Policy and Practice*, 6(1), 110-116.
- Cowie, B., Hipkins, R., Keown, P., & Boyd, S. (2011). *The shape of curriculum change*. Wellington: Ministry of Education.
- DeBoer, G. (1991). *A history of ideas in science education: Implications for practice*. New York: Teachers College Press.
- Ford, M., & Forman, E. (2006). Redefining disciplinary learning in classroom contexts. In J. Green & A. Luke (Eds.), *Rethinking learning: What counts as learning and what learning counts* (30th ed., pp. 1-32). Washington: American Educational Research Association.
- Gallas, K. (1995). *Talking their way into science: Hearing children's questions and theories, responding with curricula*. New York: Teachers College Press.
- Gilbert, J. (2005). *Catching the knowledge wave? The knowledge society and the future of education*. Wellington: NZCER Press.
- Hipkins, R., & Boyd, S. (2011). The recursive elaboration of key competencies as agents of curriculum change. *Curriculum Matters*, 7, 70-86.
- Hipkins, R., Cowie, B., Boyd, S., Keown, P., & McGee, C. (2011). *Curriculum implementation exploratory studies 2. Final report: February 2011*. Wellington: Ministry of Education.
- Millar, R. (1998). Rhetoric and reality: What practical work in science education is *really* for. In J. Wellington (Ed.), *Practical work in school science: Which way now?* (pp. 16-31). London: Routledge.
- Ministry of Education. (2007). *The New Zealand curriculum*. Wellington: Learning Media.
- Newton, P. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149-170.
- Organisation for Economic Co-operation and Development. (2005). *The definition and selection of key competencies: Executive summary*. Paris: Author. Retrieved December 2, 2005, from www.pisa.oecd.org/dataoecd/47/61/35070367.pdf.
- Resnick, L. (2010). Nested learning systems for the thinking curriculum. *Educational Researcher*, 39(3), 183-197.
- Rutherford, J. (2005). Key competencies in the New Zealand curriculum: Development through consultation. *Curriculum Matters*, 1, 210-227.
- Rychen, D., & Salganik, L. (Eds.). (2003). *Key competencies for a successful life and a well-functioning society*. Cambridge, MA: Hogrefe & Huber.
- Scalise, K., & Wilson, M. (2011). The nature of assessment systems to support effective use of evidence through technology. *E-Learning and Digital Media* 8(2), 121-132.
- Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher*, 27(2), 4-13.
- Walker, M. (2008). Capability formation and education. In B. Lingard, J. Nixon & S. Ranson (Eds.), *Transforming learning in schools and communities: The remaking of education for a cosmopolitan society* (pp. 134-151). New York: Continuum.

An Alignment of the Trends in International Mathematics and Science Survey Mathematics Scale with the 'By the End of Year 5' National Standard in Mathematics

Elliot Lawes
Ministry of Education

Email: Elliot.Lawes@nzcer.org.nz

Abstract

This paper describes an initial alignment between two broadly overlapping mathematics assessments: that used in the Trends in International Mathematics and Science Survey (TIMSS) and the 'By the End of Year 5' National Standard in Mathematics. It then uses this alignment, together with a national distribution of TIMSS scores, to produce estimates of the percentages of Year 5 students expected to be well below, below, at and above the 'By the End of Year 5' National Standard in Mathematics.

This alignment was achieved by statistically synthesising a number of component alignments – several using an adapted 'script scrutiny' methodology, and one using an adapted 'bookmarking' methodology. These component alignments each involved a panel of mathematics teacher-experts judging TIMSS items and scripts (authentic records of student performance) against the National Standards in Mathematics. The resulting collection of judgments was then statistically linked with the TIMSS measure of student mathematical proficiency.

The alignment suggests that an estimated 6% of Year 5 students are expected to be well below, 20% to be below, 33% to be at and 41% to be above the 'By the End of Year 5' National Standard in Mathematics.

Introduction

New Zealand's National Standards provide a relatively new framework for improving the performance of New Zealand's Year 1 to Year 8 students in reading, writing and mathematics, partially through improved monitoring. The National Standards have generated much debate (e.g., Courtney, 2011; Croft, 2011; Crooks, 2011) including concerns about league tables and the labelling of students. Despite this, they offer the potential for a clearer picture of student achievement and progress in reading, writing and mathematics at several levels of the schooling system than that which currently exists. The subsection of the Background section entitled 'The National Standards in Mathematics' describes some of the policy background to the National Standards as well as providing relevant details of the way performance is described by the Standards.

In contrast to the National Standards, the Trends in International Mathematics and Science Survey (TIMSS) has been used as a tool for monitoring the state of Year 5 and Year 9 mathematics (and science) in New Zealand for a number of years (Caygill & Kirkham, 2008). It provides a periodic overview of the mathematics performance of a representative sample of New Zealand's Year 5 students and contrasts this with the performance of comparable populations from a number of other countries. The subsection of the Background section entitled 'The Trends in International Mathematics and Science Survey' describes the TIMSS survey, in particular the assessment framework, the target population, and the assessment scale.

As TIMSS seeks to make international comparisons, the TIMSS assessment framework is not specifically designed to be aligned to the New Zealand Curriculum or the National Standards in Mathematics. Nevertheless, the TIMSS assessment framework 2007 (Mullis, et al., 2005), the New Zealand Curriculum (Ministry of Education, 2007) and the National Standards in Mathematics (Ministry of Education, 2009a) exhibit substantial overlap. An

issue of interest is then how performance described by the TIMSS mathematics assessment relates to performance described by the National Standards in Mathematics on the overlap between their populations of focus: Year 5 students.

TIMSS describes student performance in mathematics in a number of ways (e.g., Caygill & Kirkham, 2008), but all of these are based on a test score. This test score is generated from an item response theory model applied to data generated by students attempting a battery of items. The National Standards also allow description of performance in a number of ways (Ministry of Education, 2009d) but all of these require criterion-based reporting. Given that the description of mathematics in the TIMSS mathematics assessment is score-based and that in the National Standards it is standards-based, the relationship between these two assessments can be framed as a standard-setting issue: how do TIMSS scores align with the performance descriptors used in the National Standards in Mathematics?

A brief overview of standard-setting in the context of the current study is provided in the subsection of the Background section entitled 'Standard-setting'. The current study statistically combines the data from a number of standard-setting exercises. The details of this process are organised into two subsections of the Methodology section: one concerned with data collection entitled 'Data collection methodology for standard-setting'; and one concerned with statistical analysis entitled 'Data processing methodology for standard-setting'.

The standard-setting methodologies combined in the current work are adapted versions of 'bookmarking' (see e.g., Schagen & Bradshaw, 2003) and 'script scrutiny' (see e.g., Whetton, Twist, & Sainsbury, 2000). This combination represents one possible attempt to produce a single alignment that can be updated (by including any future standard-setting data using the same assessments), that is not overly dependent on any methodology (it currently includes two standard methodologies and could be updated to include more) and that is grounded in good standard-setting practice (the methodologies included are well documented and widely used).

The alignment itself, together with its constituent alignments and the estimates of national performance, are included in the Results section. As a quality check, the TIMSS mathematics assessment was also aligned against the 'By the end of Year 6' National Standard in mathematics. However, estimates of the percentages of Year 6 students expected to be well below, below, at and above the 'By the End of Year 6' National Standard in Mathematics were not able to be produced as there is no national distribution information for TIMSS mathematics performance for Year 6 students. This paper concludes with a discussion of the implications of some of the methodological choices made in the course of standard-setting.

Background

This section contains overviews of standard-setting, of the National Standards in Mathematics and of the TIMSS Survey.

Standard-setting

Standard-setting is broadly understood to be the task of determining a minimum test-score corresponding to some minimum competency standard, described by a set of criteria. The resulting 'cut-score' is then used to separate those test candidates who meet the minimum standard from those who do not. Hattie and Brown (2003), Näsström and Nyström (2008), and Stahl (2008), for example, provide descriptions of, and comparisons between, some of the many methodological approaches to standard-setting. Kane (2002) distinguishes between 'test-centred' and 'examinee-centred' standard-setting methods. Test-centred methods involve judgements made about the performance level on test items or tasks

required to just meet the standard. Examinee-centred methods involve judgements made about the performance level on test items *demonstrated* by examinee performance. In both test-centred and examinee-centred methods, judgements are combined to provide the required cut-score.

Both test-centred and examinee-centred methods have their benefits, drawbacks and optimally applicable contexts, and each covers a wide range of methodological approaches. The current standard-setting exercise uses a combination of two methodological approaches: script scrutiny (see e.g., Whetton et al., 2000), and bookmarking (see e.g., Schagen & Bradshaw, 2003) but adapts these somewhat.

Briefly, bookmarking involves expert judgment of items arranged according to their difficulty. Since student response is not directly incorporated, bookmarking is a test-centred standard-setting method. Script scrutiny, on the other hand, involves expert judgment of student responses to items and is therefore an examinee-centred standard-setting method.

The current adaptations to these methods attempt to account for the use of a range of evidence when judging student performance against the National Standards (Ministry of Education, 2010b). In this context, any single piece of evidence should not determine, but should only inform a judgment. Therefore, if standard-setting in this context (using one piece of evidence from a range) is to remain quantitative, it should be probabilistic rather than deterministic. In other words, a classical or deterministic approach to standard-setting associates with every score on an assessment, a single performance descriptor of a standard (with probability 1). In contrast, a probabilistic approach to standard-setting associates with every score on an assessment, each performance descriptor of the standard together with the probability of a student with that score being judged at that performance descriptor. For example, using deterministic standard-setting, a script with a score between 450 and 475 on the TIMSS mathematics assessment might be judged as at the Year 5 standard. However, using probabilistic standard-setting, the same script might be extremely unlikely to be judged well below the standard, have a 0.1 probability of being judged below the standard, have a 0.5 probability of being judged at the standard and have a 0.4 probability of being judged above the standard. Without prescribing a judgement, this might direct a teacher toward a judgement of the student who produced this script. It is evident that this probabilistic approach to standard-setting effectively abandons the idea of cut-scores altogether.

The National Standards in Mathematics

New Zealand's National Standards in reading, writing and mathematics were developed in response to the New Zealand National Party's "Crusade for Literacy and Numeracy" (National Party, 2008). The intent of this policy is to "...ensure that children get the basic skills they need to do well at school, to prepare for the workforce, and to climb the ladder of opportunity" (p. 1).

In fact, consideration of the basic skills needed to access the New Zealand Curriculum was central to the development of the National Standards: "The National Standards will help by setting clear expectations for the reading, writing and mathematics knowledge and skills students need to achieve at each level of the curriculum" (Ministry of Education, 2009b, para 3).

The Year 5 and Year 6 National Standards are most relevant to the current work. They provide descriptions of the mathematics skills expected of Year 5 and Year 6 students and are described in detail by the Ministry of Education (2009a), and then summarised (Ministry of Education, 2009e). Reporting against National Standards includes the use of four 'reporting bands': well below, below, at or above the standard for the year group, based on overall teacher judgement from a range of evidence (Ministry of Education, 2009c, NAG 2A). The Ministry of Education (2010a) provides definitions of what it means

to be judged above, at, below and well below the National Standards. In addition, the Ministry of Education, (2010a) also stipulates that a judgement of a student's achievement as: above a given standard is likely to occur if that student is achieving at a higher standard; below a given standard is likely to occur if that student is achieving at the previous year's standard; well below a given standard is likely to occur if that student is achieving at a lower level than that described by the previous year's standard. The differences in these two systems of judgment were accounted for in the standard-setting exercises.

The Trends in International Mathematics and Science Survey

The data used in this paper were a subset of New Zealand's TIMSS 2006/2007 student background data. Caygill and Kirkham (2008) provide detailed descriptions of TIMSS 2006/2007 in New Zealand. At the international level, TIMSS is described by Mullis et al. (2005), Mullis, Martin, and Foy (2008), and Olson, Martin, and Mullis (2008).

Briefly, TIMSS 2006/2007 was the third in a four-yearly cycle of international assessments of mathematics (and science) at the middle primary level. In TIMSS, the mathematics assessment framework is organised around two dimensions: a content dimension specifying subject matter to be assessed within mathematics, and a cognitive dimension specifying the thinking processes to be assessed (Mullis et al., 2005).

The content dimension was broken down into three domains: Number, Geometric Shapes and Measures, and Data Display with the greatest emphasis on Number and the least on Data Display. The cognitive dimension was also broken down into three domains: Knowing, Applying, and Reasoning. Knowing includes such behaviours as recollection, recognition, computation, retrieval, measurement, classification and ordering. Applying includes selecting, representing, modelling, implementing and solving routine problems. Reasoning includes analysis, generalisation, synthesis or integration, justification and solving non-routine problems.

The TIMSS 2006/2007 mathematics assessment consisted of a 36-minute, pencil-and-paper test containing both multiple-choice and constructed-response items. Each student received one of fourteen booklets and each booklet contained around 22 items. Each item appeared in two booklets, allowing the linking of student responses from the various booklets. The assessment was conducted in the English language and therefore was only suitable for alignment with the (English medium) National Standards. Most of the items, and consequently booklets, are not publicly released so that TIMSS can be used to measure trends. The current analysis used the four booklets containing only publicly released questions.

The TIMSS 2006/2007 target population in New Zealand was students in Year 5 in 2006. Around 4,940 Year 5 students from 220 schools participated in the main TIMSS survey. The achieved sample for New Zealand was of high quality and representative of the 2006 Year 5 student population. Olson et al. (2008) provide details of the sampling, response rates, and exclusions for New Zealand and other participating countries.

Key outputs included in the TIMSS 2006/2007 data are numeric measures of student mathematical ability and item difficulty. However, these were generated from analysis of the entire set of TIMSS items and participating students rather than the four publicly released booklets and the students who responded to them. Measures of mathematical ability and item difficulty based on the latter were generated using a Rasch model. These measures of mathematical ability had an extremely high correlation (r -squared greater than 0.95) with those based on the entire TIMSS sample. For those already familiar with TIMSS international reporting, the Rasch score was re-standardised to have mean 485.7 and standard deviation 85.3 – the mean and standard deviation of respondents to the four publicly released TIMSS booklets measured on the international TIMSS reporting scale. The item difficulty measures were similarly re-standardised.

Methodology

This section contains descriptions of the methodology used in the current study. It is broken into two subsections – the first dealing with data collection and the second dealing with statistical analysis of the collected data.

Data Collection Methodology for Standard-setting

In total, seven standard-setting exercises were carried out in order to collect data to align the TIMSS mathematics scale to the National Standards in Mathematics. Table 1 summarises these. The descriptors in Table 1 are explained in the following text.

Table 1: Data collection for standard-setting

ID	Standard-setting methodology	Judgment System	Standard	Materials	Judgments
1	Bookmark	Best-fit	All	Item booklet	2684
2	Script scrutiny	Best-fit	All	Script set 1	90
3	Script scrutiny	Best-fit	All	Script set 2	90
4	Script scrutiny	Four-point	Year 5	Script set 1	90
5	Script scrutiny	Four-point	Year 6	Script set 1	90
6	Script scrutiny	Four-point	Year 5	Script set 2	90
7	Script scrutiny	Four-point	Year 6	Script set 2	90

The 'ID' column enables matching of information in Table 1 with that in Tables 2 and 3. As mentioned above, two standard-setting data collection methodologies were used – the bookmark and script scrutiny – but these were adapted somewhat. Both used the same panel of fifteen experts as judges. These panellists were experienced numeracy professional development facilitators and senior teachers selected because of their familiarity with New Zealand's curriculum and, in particular, with the National Standards documents.

In the case of the bookmarking methodology, each panel member independently placed bookmarks throughout a booklet of items which had been ordered by difficulty. The bookmarks separated those items likely to be correctly answered by a minimally competent student at a given level of the National Standards from the remaining items. This was followed by a discussion of the panel members' decisions and further rounds of bookmarking and discussion. In the more usual 'deterministic' bookmarking (Schagen & Bradshaw, 2003) the further rounds of bookmarking and discussion are directed to attain a consensus of decision. In the current 'probabilistic' bookmarking, there was no explicit attempt to attain consensus of decision. Rather, all discussions were focussed on consistent use of evidence and interpretation of the standards.

In the case of the script scrutiny methodology, following a practice session and discussion, each panel member independently judged a selection of student scripts (completed TIMSS tests) against the standards. In the more usual 'deterministic' script scrutiny (Whetton et al., 2000) cut-scores are determined during analysis of the collected data. In the current 'probabilistic' script scrutiny, the same data collection procedure was used as in the 'deterministic' script scrutiny but a different post-data-collection analytical procedure was employed.

Two judgment systems were used: Best-fit and Four-point. The Best-fit judgement system refers to panellists judging which standard best fits either student work or testing materials. For example, a panellist might compare a piece of student work with the standards and decide that the standard that best described the work was the Year 6 Standard. The four-point judgement system refers to panellists judging whether student work or testing

materials is well below, below, at or above a pre-specified standard. For example, a panellist might compare a piece of student work with the Year 5 Standard and decide that the work was above the Year 5 Standard. The equating of these two systems prescribed by the Ministry of Education (2010a) is described above in the section The National Standards in Mathematics.

As described above, the best-fit judgment system required the panellists to use all of the standards as comparators. However, use of the four-point judgment system required either the Year 5 or Year 6 Standard to be specified prior to standard setting. The focus of the alignment was on the Year 5 Standard, but as a quality control measure, and to provide the panellists with a different perspective on their judgments, data to align the TIMSS mathematics scale with the Year 6 Standard were also collected.

Three sets of materials were used. For the bookmarking exercise, the materials consisted of a booklet of all of the sixty items from TIMSS booklets one to four. For the script scrutiny exercises, two non-overlapping sets of scripts were used. Each set consisted of 30 distinct scripts arranged into 15 packs of six scripts each – one pack for each panellist. Each script was judged by three panellists. The thirty scripts in both sets were made up of three scripts randomly selected from each TIMSS mathematics scale performance decile, and therefore represented the full range of performance.

For the bookmarking exercise, fifteen judges, each making three rounds of judgments of sixty judgments (or items) per round resulted in a possible 2700 judgments. One judge was absent for one of the rounds and there was a single data entry error leaving 2684 usable records. For each script scrutiny exercise, each of 30 scripts was judged three times resulting in 90 usable records.

Data Processing Methodology for Standard-setting

There were two major methodological factors to consider in data processing. The first of these was the need to take a probabilistic perspective when reporting the findings of the standard-setting exercises. The second of these was the need to compare and synthesise the data collected in each of the seven different standard-setting exercises described in the previous section.

As mentioned above, the probabilistic approach to standard setting in the context of the National Standards is motivated by teachers' use of multiple sources of evidence when making a judgment. As the collection of standard-setting data was designed with this in mind, all recorded judgments were simply and statistically summarised using a multinomial logistic regression model. This model provides the probabilities of a student with a given TIMSS mathematics score being well below, below, at or above the appropriate National Standard.

This methodology was applied to the data from each of the standard-setting exercises described in the previous section (excluding those judged against the Year 6 Standard). The resulting model parameter estimates are provided in Table 2. The alignments resulting from the script scrutiny datasets were reasonably similar. However, each of these alignments was quite different to those resulting from the bookmarking dataset (also see Table 2). In particular, the bookmarking data indicated that no TIMSS items were well below the Year 5 Standard whereas the script scrutiny data contained scripts that were judged well below the Year 5 Standard. Such scripts typically contained a few sporadically distributed correct answers amongst many unattempted items and incorrect answers. That the bookmarking exercise revealed no well below items is perhaps not surprising given the aims, scope and development of TIMSS. However, as the current work represents the first quantitative standard-setting relating TIMSS to New Zealand's curriculum, the possibility of well below items could not be ruled out *a priori*.

Table 2: Model Parameter Estimates

ID	Year-level	Well below/Below (se)	Below/At (se)	At/Above (se)	'Slope' (se)
1 (a)	Year 5	.	-5.69 (0.24)	1.58 (0.12)	-9.56 (0.38)
1 (b)	Year 6	-6.09 (0.23)	1.70 (0.12)	8.31 (0.29)	-10.30 (0.34)
2 (a)	Year 5	-1.64 (0.42)	0.33 (0.37)	3.02 (0.51)	-2.42 (0.35)
2 (b)	Year 6	0.39 (0.40)	3.09 (0.58)	5.10 (0.78)	-2.54 (0.41)
3 (a)	Year 5	-3.09 (0.59)	0.31 (0.40)	3.32 (0.61)	-3.02 (0.44)
3 (b)	Year 6	0.20 (0.36)	2.62 (0.52)	4.34 (0.68)	-2.32 (0.37)
4	Year 5	-2.35 (0.49)	0.56 (0.38)	3.30 (0.54)	-2.57 (0.36)
5	Year 6	-0.74 (0.42)	2.42 (0.52)	5.60 (0.85)	-2.88 (0.44)
6	Year 5	-3.73 (0.64)	-0.84 (0.42)	3.59 (0.65)	-2.99 (0.44)
7	Year 6	-1.31 (0.38)	1.48 (0.39)	4.77 (0.69)	0.32 (0.32)
Weighted (a)	Year 5	-3.76 (0.21)	-1.17 (0.13)	1.32 (0.12)	-2.49 (0.13)
Weighted (b)	Year 6	-1.43 (0.13)	1.18 (0.13)	3.75 (0.20)	-2.74 (0.14)

This phenomenon impacted on the synthesis of the alignments. In particular, because the bookmarking alignment included no probability distribution for the well below judgment, no straightforward averaging of the probability distributions from different alignments was possible (as was implemented by the Ministry of Education, 2011). Instead, a data-weighting approach was taken so that the bookmarking data (of which there were more) contributed the same total weight to the logistic alignment model as the combined script scrutiny data (of which there were less). The model parameters for this 'weighted' model are provided in Table 2.

It should be noted that the TIMSS mathematics framework has remained independent of the curricula of any of the countries participating in TIMSS. Also, significant effort has been expended ensuring that the TIMSS mathematics scale remains unchanged over time. Therefore, the current methodology did not have to account for any recent changes in New Zealand's curriculum.

Results

This section contains numerical and graphical summaries of the standard-setting exercise. Table 2 contains estimates of the model parameters from the overall alignment as well as all of the constituent alignments.

The ID column enables matching of information in Table 2 with that in Tables 1 and 3. For example, ID 1(a) can be matched with ID 1 in Table 1, as can ID 1(b). The presence of an '(a)' in the ID column indicates that data were collected using the best-fit judgment system (using all standards as comparators) and then converted to a judgment against the Year 5 Standard. The presence of a '(b)' in the ID column indicates that data were collected using the best-fit judgment system (using all standards as comparators) and then converted to a judgment against the Year 6 Standard. This conversion is described in the subsection entitled The National Standards in Mathematics. The entries in the Year-level column indicate the year-level of the students being modelled. The entries in the remaining columns are regression coefficients and their standard errors.

To understand Table 2, note that the data in the Weighted (a) row could be interpreted as follows: the probability, $WB(x)$, of a TIMSS assessment with a score of x being judged well below the Year 5 National Standard in Mathematics is approximately:

$$WB(x) = \frac{e^{-3.76-2.49x}}{1 + e^{-3.76-2.49x}}$$

The probability, $BE(x)$, of the same script being judged below this standard is approximately:

$$BE(x) = \frac{e^{-1.17-2.49x}}{1 + e^{-1.17-2.49x}} - WB(x)$$

The probability, $AT(x)$, of the same script being judged at this standard is approximately:

$$AT(x) = \frac{e^{1.32-2.49x}}{1 + e^{1.32-2.49x}} - WB(x) - BE(x)$$

Finally, the probability, $AB(x)$, of the same script being judged above this standard is approximately:

$$AB(x) = 1 - WB(x) - BE(x) - AT(x)$$

The figures in parentheses are estimates of the standard errors of the parameters.

Figure 1 depicts the probability distributions from the Weighted (a) and Weighted (b) models (i.e., the synthesised alignments against the Year 5 and Year 6 Standards).

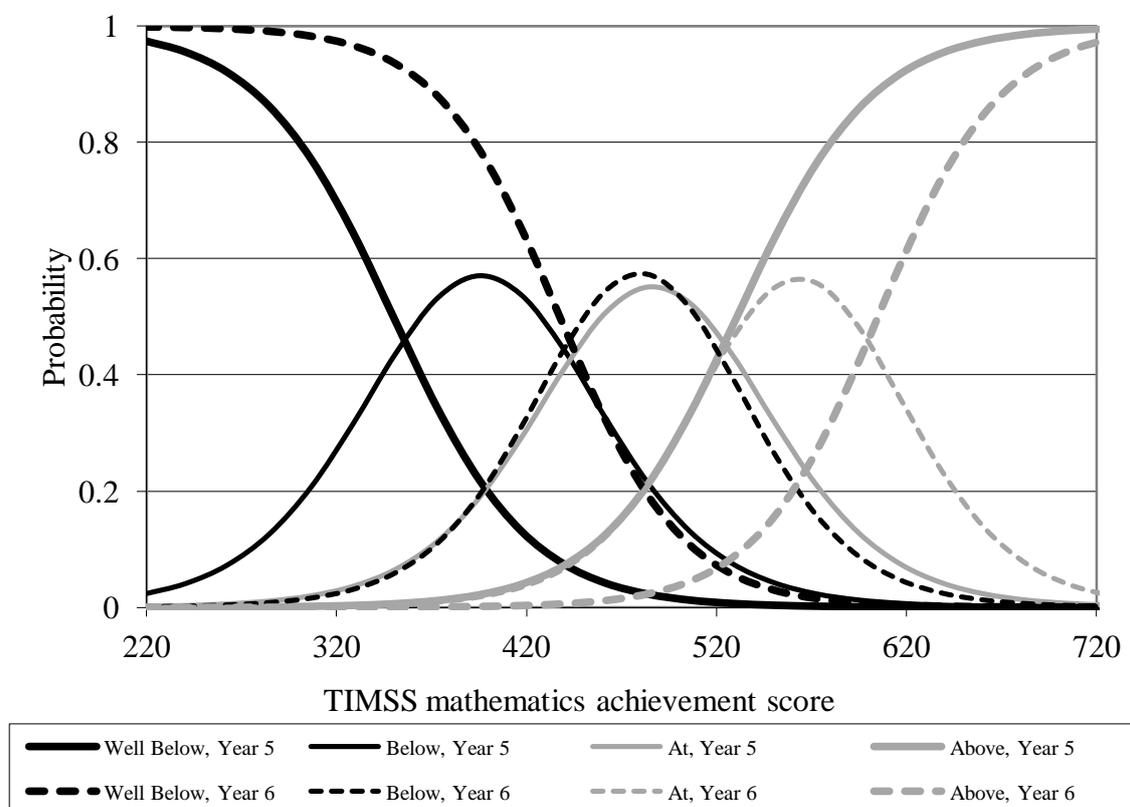


Figure 1: The National Standards in Mathematics and the TIMSS Mathematics Scale

As might be expected, the probability curves for the alignment against the 'By the end of Year 6' Standard are to the right of the analogous curves for the alignment against the 'By the end of Year 5' Standard. This indicates that the judges found the 'By the end of Year 6' Standard more difficult than the 'By the end of Year 5' Standard.

Figure 2 indicates the method used to estimate the percentages of Year 5 students expected to be well below, below, at and above the 'By the End of Year 5' National Standard in Mathematics. Explicit knowledge of the national distribution of TIMSS performance was a key feature of the production of these estimates. It uses the weighted (a) alignment (i.e., the synthesised alignment against the Year 5 Standard). To compute these estimates, the areas in Figure 2 of the same colour were added.

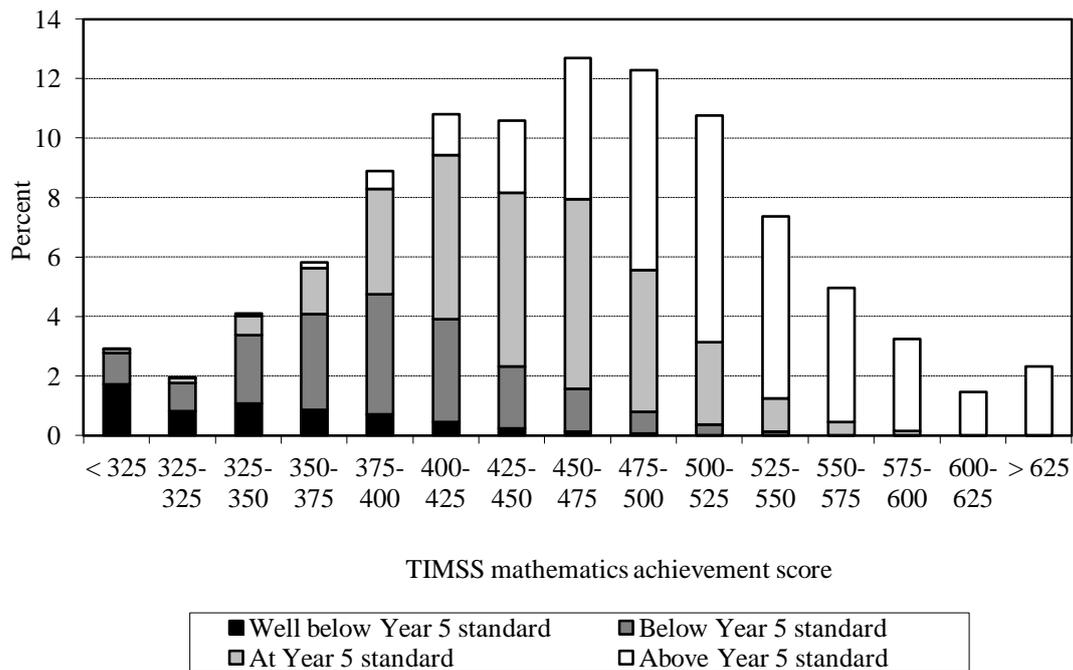


Figure 2: The TIMSS Mathematics Score Distribution and the 'By the end of Year 5' National Standard in Mathematics

Table 3 shows, for each of the constituent alignments and the synthesised weighted (a) alignment, estimates of the percentages of Year 5 students expected to be well below, below, at and above the 'By the End of Year 5' National Standard in Mathematics. All of the estimates in Table 3 were produced using a method similar to that depicted in Figure 2. As these estimates are quite crude, no estimates of standard error are included.

Table 3: Estimated Distributions of National Standards Reporting Bands by Alignment

ID	Well below	Below	At	Above
1 (a)		17	27	55
2 (a)	21	25	35	20
3 (a)	12	32	34	22
4	15	33	34	18
6	8	22	50	19
Weighted (a)	6	20	33	41

Table 3 demonstrates that the script scrutiny alignments (ID 1, 2 (a), 3 (a), 4 and 6) were somewhat different to the bookmarking alignment (ID 1 (a)). The weighted (a) estimates capture features of both the script scrutiny and bookmarking estimates.

The TIMSS 2006/07 international benchmarks of mathematics achievement – themselves a type of cut-score – are described in detail in Mullis et al. (Exhibit 2.1, 2008). Briefly, students at or above the *Advanced* international benchmark (625) can apply their understanding and knowledge in a variety of relatively complex situations and contexts and can explain their reasoning. For example, they can apply proportional reasoning, select appropriate information to solve multi-step word problems, and apply geometric knowledge of a range of two- and three-dimensional shapes.

Students at or above the *High* international benchmark (550) can apply their knowledge and understanding to solve problems. For example, they can solve multi-step word problems involving operations with whole numbers, use division in a variety of situations, and show some basic geometric knowledge. Students at or above the *Intermediate* international benchmark (475) can apply basic mathematical knowledge in straightforward situations. For example, they understand whole numbers, they can extend simple numeric and geometric patterns and are familiar with a range of two-dimensional shapes. Students at or above the *Low* international benchmark (400) have basic mathematical knowledge. For example, they have an understanding of adding and subtracting with whole numbers and familiarity with triangles and informal coordinate systems.

Table 4, obtained from the TIMSS mathematics score distribution, the TIMSS International Benchmarks and the alignment shows the percentages of Year 5 students in each TIMSS benchmark and National Standard reporting category. To read the table, note that the '2' in the cell indexed by Well below and Low-Intermediate indicates that the alignment estimates that approximately 2% of Year 5 students would be judged well below the Year 5 Standard in mathematics *and* would have TIMSS mathematics scores between 400 (Low) and 475 (Intermediate). As noted previously, the TIMSS mathematics score distribution reported here is based on a Rasch ability score generated using TIMSS booklets one to four which has then been re-standardised to have mean 485.7 and standard deviation 85.3. This distribution is slightly different to that used for international TIMSS reporting (as used in, for example, Mullis et al., 2008). As for Table 3, the estimates in Table 4 are quite crude, so no estimates of standard error are included.

Table 4: National Standards and the TIMSS International Benchmarks

TIMSS International Benchmark category	'By the end of Year 5' National Standard category				
	Well below	Below	At	Above	Sum
Below 'Low'	4	4	1	0	9
Low' - 'Intermediate'	2	11	11	2	25
Intermediate' - 'High'	0	4	17	14	36
High' - 'Advanced'	0	0	4	18	23
Above 'Advanced'	0	0	0	7	7
Sum	6	20	33	41	100

Table 4 shows that:

- the Well below the 'By the End of Year 5' National Standards aligned mostly with the Below Low TIMSS mathematics performance descriptor
- the Below the 'By the End of Year 5' National Standards aligned mostly with the Low-Intermediate TIMSS mathematics performance descriptor
- the At the 'By the End of Year 5' National Standards aligned roughly equally with the Low-Intermediate and Intermediate-High TIMSS mathematics performance descriptors
- the Above the 'By the End of Year 5' National Standards aligned roughly equally with the Intermediate-High and High-Advanced TIMSS mathematics performance descriptors.

Discussion

The present work is an initial attempt to quantitatively align mathematical performance against the National Standards with that against an international assessment framework. As such, it is exploratory in nature and can only offer tentative conclusions. However, it does raise some interesting questions. Three of these are:

1. What are appropriate and robust ways to use expert judgment and standard-setting to estimate the percentages of students expected to be at the various reporting bands of the National Standards?
2. How might estimates from Question 1 be compared with teacher judgement against the National Standards?
3. If a comparison as in Question 2 finds unacceptably large differences between expert-made and teacher-made judgments against the National Standards, what tools might be used to decide the extent to which the standards themselves are not set appropriately?

These are discussed in order.

Centralised collection of data does not yet feature in the design of the National Standards, and it is therefore likely that expert judgement and estimation will be necessary in order to gauge how many students are expected to be at the various reporting bands. The current work presents one approach to answering Question 1, but it only uses a single assessment tool – the TIMSS mathematics assessment – among several possibilities (for example PAT: Mathematics and e-asTTle mathematics). How estimates from different assessments (each with its own framework and relationship with the National Standards) will be compared and perhaps synthesised is surely a complex and expensive issue which is likely to change over time.

The question of whether or not to synthesise is itself quite gnarly. Synthesis enables the production of a set of estimates which are likely to be robust to changes over time but which are methodologically opaque and difficult to explain. Non-synthesis necessitates the production of many, perhaps disparate and unstable, sets of estimates, each of which is methodologically transparent. Both approaches have their benefits and drawbacks.

After the National Standards have become embedded in teacher practice, they, and their use, will need to be periodically evaluated. If any such evaluation is to be robust, then as Question 2 suggests, 'acceptable' differences between expected and realised percentages of students at the various reporting bands will need to be explicitly stated. But when estimation procedures are as crude as those in the current work, error in estimation – and therefore in significance of difference between estimates – is difficult to quantify meaningfully. This again emphasises the need for robust and appropriate estimation methods and also indicates that in order to minimise error, consideration may need to be given to fairly large-scale standard-setting work.

If the comparisons described in Question 2 reveal 'unacceptable' differences between expected and realised percentages of students at the various reporting bands, then the next likely question is 'why?' That is, as Question 3 asks, does the cause of the mismatch between expected and attained percentages lie mostly with students and learning, with teachers and instruction, with the standards themselves and judgment or some combination of these and other factors?

Internationally comparative studies such as TIMSS surely offer an insight into this. For example, if the expert judgment and standard setting produce estimates of percentages 'above' and at the appropriate National Standards that were lower than any estimates obtained from teacher judgements, it may not be clear whether or not experts and the standards themselves are too harsh or teachers too lenient. However, if New Zealand students compared well internationally, then this might lend credence to the possibility that the National Standards were too difficult. Understanding of the relationship between the TIMSS mathematics assessment framework and that of the National Standards then becomes paramount.

References

- Caygill, R., & Kirkham, S. (2008). *Mathematics: Trends in year 5 mathematics achievement 1994 to 2006*. Wellington: Ministry of Education.
- Courtney, B. (2011). National Standards and the accountability question. *New Zealand Principal*, 26(1), 14-15.
- Croft, C. (2011). National Standards and league tables: Political agenda. *New Zealand Principal*, 26(1), 11-13.
- Crooks, T. (2011). National Standards revisited. *New Zealand Principal*, 26(1), 8-9.
- Hattie, J. A., & Brown, G. T. L. (2003, August). *Standard-setting for asTTle reading: A comparison of methods*. *asTTle Technical Report #21*. University of Auckland and Ministry of Education. Retrieved November 16, 2009, from <http://www.tki.org.nz/r/asittle/pdf/technical-reports/techreport21.pdf>
- Kane, M. (2002). Conducting examinee-centred standard-setting studies based on standards of practice. *The Bar Examiner*, 71(4), 6-13. Retrieved December 21, 2009, from http://www.ncbex.org/uploads/user_docrepos/710402_kane.pdf
- Ministry of Education. (2007). *The New Zealand curriculum for English-medium teaching and learning in years 1-13*. Wellington: Learning Media.
- Ministry of Education. (2009a). *Mathematics standards for years 1-8*. Wellington: Ministry of Education.
- Ministry of Education. (2009b). *Fact sheet 6: National Standards and the NZC*. Retrieved December 21, 2009, from <http://nzcurriculum.tki.org.nz/National-Standards/Key-information/Fact-sheets/National-Standards-and-the-NZC>
- Ministry of Education. (2009c). *The national administration guidelines (NAGs)*. Retrieved August 9, 2010, from <http://www.minedu.govt.nz/NZEducation/EducationPolicies/Schools/PolicyAndStrategy/PlanningReportingRelevantLegislationNEGSAndNAGS/TheNationalAdministrationGuidelinesNAGs.aspx>
- Ministry of Education. (2009d). *Fact sheet 12: National administration guidelines (NAGs)*. Retrieved May 27, 2011, from <http://nzcurriculum.tki.org.nz/National-Standards/Key-information/Fact-sheets/NAGs-fact-sheet>
- Ministry of Education. (2009e). *The New Zealand curriculum mathematics standards for Years 1 - 8*. Retrieved May 27, 2011, from <http://nzcurriculum.tki.org.nz/National-Standards/Mathematics-standards/Supporting-resources>
- Ministry of Education. (2010a). *Definitions of achievement*. Retrieved March 30, 2010, from <http://assessment.tki.org.nz/Overall-teacher-judgment/Definitions-of-achievement>
- Ministry of Education. (2010b). *Making an overall teacher judgment*. Retrieved June 7, 2011, from <http://assessment.tki.org.nz/Overall-teacher-judgment/Making-an-overall-teacher-judgment>

- Ministry of Education. (2011). *PAT mathematics*. Retrieved June 7, 2011, from <http://assessment.tki.org.nz/Assessment-tools-resources/Alignment-of-assessment-tools-with-National-Standards/Mathematics/PAT>
- Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., Arora, A. & Erberber, E. (2005). *TIMSS 2007 assessment frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Näsström, G., & Nyström, P. (2008). A comparison of two different methods for setting performance standards for a test with constructed-response items. *Practical Assessment, Research & Evaluation*, 13(9), 1-11. Retrieved November 16, 2009, from: <http://pareonline.net/pdf/v13n9.pdf>
- National Party. (2008). *Education policy: Crusade for literacy and numeracy*. Retrieved December 21, 2009, from http://www.national.org.nz/files/2008/Education_Policy_Crusade_for_Literacy.pdf
- Olson, J. F., Martin, M. O., & Mullis, I. V. S. (Eds.). (2008). *TIMSS 2007 technical report*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Schagen, I., & Bradshaw, J. (2003, October). Use of bookmark for setting standards in reading tests. *Paper presented at the 29th IAEA conference, Manchester*.
- Stahl, J. (2008, September). *Standard-setting methodologies: Strengths and weaknesses*. Paper presented at IAEA Conference, Cambridge. Retrieved December 15, 2010, from http://www.iaea2008.cambridgeassessment.org.uk/ca/digitalAssets/180502_Stahl.pdf
- Whetton, C., Twist, E., & Sainsbury, M. (2000, April). National tests and target setting: Maintaining consistent standards. *Paper presented at AERA Annual Meeting, New Orleans*.

Beyond Assessment: Assuring Student Learning in Higher Education

Victoria J. Mabin and Stephen J. Marshall
Victoria University of Wellington

Email: vicky.mabin@vuw.ac.nz; stephen.marshall@vuw.ac.nz

Abstract

Setting up an 'Assurance of Learning' (AoL) system in line with requirements for accreditation is generally perceived to be a challenging task in both theory and practice. This paper provides an overview of the AoL system developed by the Faculty of Commerce and Administration to meet the requirements for accreditation by the Association to Advance Collegiate Schools of Business (AACSB), and describes its rationale, results achieved to date, and current challenges. The Faculty's system draws on the use of graduate attributes (Barrie, 2004), constructive alignment (Biggs, 1999), quality systems (Deming, 1982) and Theory of Constraints (Goldratt, 1994). In particular, individual student assessment is used to provide programme-level assurance of learning of graduate attributes. AoL's focus on 'closing the loop' – using student cohort performance data to inform system level change so that more students achieve the overall programme-level learning goals – is illustrated through a number of examples. While AoL developments have been led largely by business schools, we argue that wider adoption would allow universities to back up their claims about their students' achievement of graduate attributes, moving towards assuring, not just assessing, student learning.

Keywords: Accreditation, AACSB, assurance of learning, assessment, graduate attributes

Introduction

An Assurance of Learning (AoL) system has been implemented at Victoria University's Faculty of Commerce and Administration (the Faculty) as a quality-assurance approach to improving teaching and learning outcomes, and to gain accreditation from the international Association to Advance Collegiate Schools of Business (AACSB, 2007). To gain accreditation, an institution must, *inter alia*, demonstrate it meets AACSB's set of AoL Standards, which require processes aimed at continuous improvement at the programme (i.e., degree) level. However, these standards are not prescriptive and it is up to each institution to develop its own AoL system, tailored to its own mission while meeting AACSB's AoL Standards.

Setting up an AoL system is generally regarded as a challenging task, both in theory and in practice, with many business schools struggling to understand what is required as well as how it can be implemented (Martell, 2009; E. Peacock, personal communication, March, 9, 2011). Our system has been successfully established, drawing on management theories and quality improvement processes as well as higher education literatures, resulting in a system that is feasible and has led to favourable evaluations from external evaluators. Victoria's system, which has so far involved more than 250 assessment exercises conducted in undergraduate and postgraduate taught courses, as well as thesis examination reports, has been described as being achieved in record time, and thus may provide pointers to other business schools.

As one measure of 'success', formal accreditation was granted to the Faculty by AACSB in 2010. Continuing improvements to programmes and processes have produced other valuable results. Internal recognition of the benefits of the AoL processes is evidenced by the adoption of AoL by programmes 'outside the scope' of AACSB accreditation requirements. Related outcomes include meaningful changes to course activities, support and assessment systems, with widespread adoption of assessment criteria such as rubrics, and observable

improvement in student learning outcomes. The overall project has seen a positive shift in the way the Faculty engages with programme and course design, and with teaching and learning as a whole. It has motivated a stronger pedagogical process at programme and course level, guiding changes in student and teacher support structures and focus. As a result of the AoL process, the Faculty is able to gauge the levels of achievement of programme learning goals, and monitor the consequences of modifying course(s) and programmes. 'Closing the loop' in the AoL cycle provides a way of working towards assuring that our teaching and learning objectives are met.

Adoption of such processes more widely – beyond business schools – may provide an avenue for universities that have adopted graduate attributes to monitor and progressively improve student achievement of such attributes. The AoL process is structured by the use of programme learning goals aligned through curricula and assessed directly, in order to enhance the achievement of learning goals by the student cohort. The guiding philosophy is to use student cohort performance data to inform system level change so that students achieve the overall programme-level learning goals, in contrast to the more normal assessment where the focus is on helping individual students perform better in individual courses.

Banta (2008) advocates the use of assessment data beyond individual courses, to address programme and institutional level concerns in much the same way that AACSB envisages the use of data to improve programme effectiveness. Interestingly, Banta claims that reports in the literature have focused on improvements to processes and practices, but have provided little evidence that such activities have been effective in improving student learning outcomes. While it is acknowledged that more evidence needs to be gathered and analysed, this paper provides some early evidence of improvement in learning outcomes through adoption of the AoL process.

After a brief literature review, the paper describes the AoL process developed, relating it to the higher education literature in particular on graduate attributes. Significant features of this system and examples of results are provided, before discussing challenges currently being addressed. Finally the relevance for non-business subjects is discussed.

Perspectives from the Literature

The AoL system is anchored explicitly on the concept of graduate attributes, which are encapsulated and contextualised in programme learning goals. The underlying belief is that universities should be able to describe what their graduates will be capable of, and how the education they receive will support their development and attainment of those capabilities. Furthermore, they should be able to demonstrate that all their graduates have achieved these graduate attributes.

Graduate attributes are not a new concept to higher education with some arguing that they date back to at least 1862 and the establishment of the University of Sydney (Barrie, 2004). In the modern sense they are “the qualities, skills and understandings a university community agrees its students would desirably develop during their time at the institution and, consequently, shape the contribution they are able to make to their profession and as a citizen” (Bowden, Hart, King, Trigwell, & Watts, 2000). The concept, either explicitly or implicitly, includes the idea that the attributes will encompass disciplinary skills and knowledge as well as generic attributes: “These are the skills, personal attributes and values which should be acquired by all graduates regardless of their discipline or field of study. In other words, they should represent the central achievements of higher education as a process” (Higher Education Council, 1992, p. 20).

Over the last two decades graduate attributes have also become a key instrument in the accountability frameworks of both the United Kingdom (as a result of the Dearing Report; NCIHE, 1997) and Australia through the Australian Universities Quality Agency (AUQA,

now the Tertiary Education Quality and Standards Agency, TEQSA). In the context of commerce education, graduate attributes are a key component of the systems required for both AACSB and EQUIS (European Quality Improvement System) accreditation. Despite, or perhaps because of, this history of activity there is now a confusion of terminology in the field with graduate attributes framed as transferable, key, core, generic, learning, lifelong, or personal, and the attributes re-expressed as skills, goals, competencies or capabilities (Clanchy & Ballard, 1995). In many cases, but not all, these are used interchangeably but they can be intended to make fine distinctions. This confusion in part represents a symptom of the challenge that faces many institutions when defining and operationalising attributes into actions and educational choices.

Research on academic perceptions of what graduate attributes should encompass has shown a significant diversity of conceptions (Barrie, 2004). This range of conceptions means that inevitably graduate attributes need to be expressed at a high level of abstraction and then contextualised, commonly through curriculum mapping into specific qualifications, programmes or courses of study. This mapping process now generally sees the graduate attributes expressed and translated into a subset of the learning objectives of courses and then, through a process of constructive alignment (Biggs, 2003) mapped to specific learning activities and assessments. Constructive alignment emphasises that it is necessary to decide on the goals of a programme, and to develop these skills, knowledge, and disposition throughout the programme in a deliberate and planned way, integrated with assessment. Such an approach might be seen as an 'interlocked' approach, as defined by the Globalization of Management Education Task Force (2011). They argue for planned development of skills, stating that practices of infusion (where everyone is supposedly building skills) and insertion (where a few isolated exercises are included in the curriculum) do not work. They argue that infusion leads to invisibility while insertion leads to isolation, and what is needed is a more connected interlocked approach.

There are many descriptions of the experience of institutions engaging in the process of mapping graduate attributes for programme curriculum (re)design, many of which note that the process can be less than straightforward. Treleaven and Voola (2008) describe the process of integrating graduate attributes in Marketing education at the University of Sydney and noted significant differences when a constructive alignment approach was taken to embed the graduate attributes into courses. The experience of the Open University UK (Dillon, Reuben, Coats, & Hodgkinson, 2007) emphasised the importance of integration of assessment and the need to ensure staff are appropriately supported as factors necessary for the achievement of a common understanding of the attributes and ultimately a successful outcome for students.

Sumsion and Goodfellow (2004) report a generally positive outcome from mapping graduate attributes but noted the need to address staff concerns about the process leading to an overly managerial and audit-focused culture, something of particular concern when developing an AoL process in the context of an external accreditation activity. Sumsion and Goodfellow also noted that care should be taken to ensure that the semantic meaning of the maps is collectively understood and that maps are made with a common set of standards and expectations. This latter issue has also been identified by Crebert (2002), particularly when mapping a standard set of attributes across multiple disciplines. De La Harpe, Radloff, and Wyber (2000) noted issues with staff resenting a 'top-down' approach and lacking commitment and time to engage in the project, including in the professional development and support provided.

Despite these issues, there has been widespread adoption of graduate attributes by Australasian institutions. Such is the level of activity that there has been a series of substantial projects funded in Australia looking at the ways graduate attributes are created, operationalised and outcomes measured, and the type of graduate attributes in use throughout Australian universities (The National GAP, 2011; Oliver, 2011) in an attempt to understand the extent and diversity of ways graduate attributes are influencing learning and teaching. It is apparent that while graduate attributes are commonly stated in

university materials, and such attributes are implicitly covered in programmes of study, few institutions would be able to demonstrate that they are actually achieved by all graduates. Many universities would find it difficult even to ascertain what proportion of their graduates has actually achieved them. The AoL process is designed to quantify and assure the claims made concerning graduate attributes, and would provide universities with a mechanism for backing up their claims.

Aside from the measurement of achievement of graduate attributes, it is not uncommon that an organisation (or discipline) can find it difficult to improve itself, including making changes that lead to improvements in student learning outcomes. Deming argued that, “competent men (sic) in every position, if they are doing their best, know all there is to know about their work except how to improve it. Help towards improvement can only come from some other kind of knowledge. Help may come from outside the company, combined with knowledge already possessed by people within the company but not being utilized” (Deming, 1982, p.143). One of Deming’s 14 Points, “Cease dependence on mass inspection,” advocating the use of quality assurance not quality control, is particularly relevant in our context, with its focus on process improvement, as opposed to 100% inspection at the end. Rather than expecting improvements to result from checking every student’s performance (assessing each item to decide on an individual’s grade), the AoL process relies on checking a sample of students’ work over a period of time to detect patterns of performance so as to intervene with system improvements to improve student learning outcomes.

Deming’s ideas are often credited as stimulating the development of quality management approaches such as Total Quality Management (TQM) which has driven substantial quality initiatives in many industries. The success of historical application of TQM to higher education is, however, contested (Houston, 2007; Koch, 2003). Koch’s review of quality management’s impact on higher education notes that the common failing of many initiatives has been to focus on aspects of higher education that are peripheral to the major purpose of supporting student learning. Houston provides an extensive critique of the problems of applying standard quality models such as TQM to higher education, particularly those aspects that are not essentially administrative. He argues for the need to take a critical systems approach, starting with clearly stated objectives for the educational activities being improved: “authentic quality theory is essentially systemic; attending to values, purpose, and optimising performance relative to the aim of the system. It offers a framework of ideas to help improve future performance, based on data about past and current performance” (p. 13).

There is also the need to address the conflict between the sense that quality management is management *by* quality rather than management *for* quality (Connor, 1997), which is inconsistent with the autonomous nature of academic work and perhaps lies at the heart of the resistance and distrust of quality mechanisms encountered in higher education (Hoecht, 2006). Perfunctory compliance can be an undesired outcome of systems that have been ill-thought through or mandated from above in a heavy-handed way (Davies, 2010).

Despite these issues and previous failed TQM initiatives, De Jager and Nieuwenhuis (2005) suggest that the underlying principles of TQM are consistent with those that underlie constructive alignment and an outcomes-driven approach to pedagogical design and practice. The AoL approach presented in this paper reflects an understanding of the need to develop applications of the ideas underlying TQM in manners that support academic engagement with pedagogical decision making. Outcomes-driven analysis, combined with the evidence gathered in the AoL process, drives the continuous improvement focus that lies at the heart of TQM.

The various issues raised above underline the difficulties of setting and achieving standards not just from the technical – in this case, pedagogical – viewpoint but also the political and economic feasibility of change. Several management theories provide frameworks to guide

change management, including the Theory of Constraints which provides a process to guide the design and implementation of change (Goldratt, 1994; Mabin, Forgeson, & Green, 2001). The development and implementation of the AoL System have been viewed by the first author as an action research project, in which the Theory of Constraints methodology has been used to identify key constraints or obstacles and address those in turn. Mabin (2009) describes this overarching framework for the development and implementation of the AoL system addressing the political and economic aspects, while in this present paper, we concentrate on the pedagogical aspects.

Description of the Approach Taken and Process Developed

In this section we will provide an overview of the AoL system that the Faculty has put into place drawing on the aforementioned literatures, while the next section will discuss the results we have achieved and take the opportunity to reflect on our approach, comparing it with other systems, and to comment on current challenges.

The Faculty has sought accreditation from various international bodies, including the AACSB. To achieve AACSB accreditation, it is necessary to provide evidence that a continuous improvement process is in place that is focused on improving the achievement of student learning outcomes. The approach can be encapsulated in the following four questions:

1. What do we want our graduates to know, do, and be?
2. Where/how will we provide opportunities for them to learn these things?
3. How will we know if the students have acquired them?
4. What will **we** (the Faculty) change if the students haven't learnt these things?

These form a continuous improvement cycle, akin to a Plan-Do-Check-Act cycle, inspired by Shewhart but popularised by Deming (1982, p. 88), as shown in Figure 1.

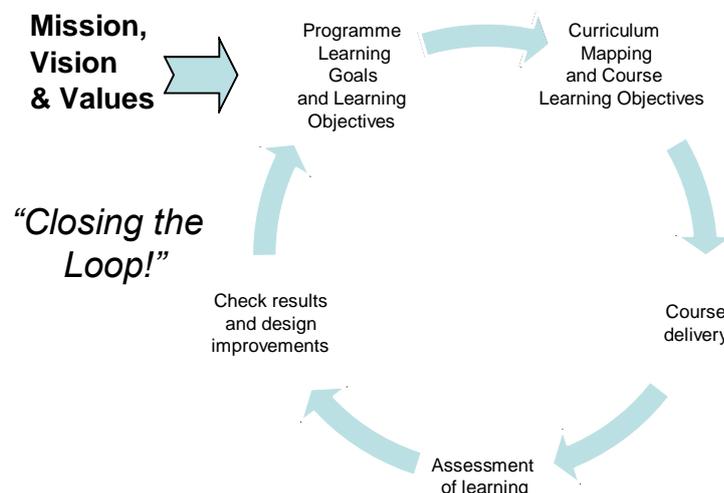


Figure 1: The Assurance of Learning Cycle

'What we want our students to know' is guided by our mission, vision, and values, which in turn inform the learning goals and learning objectives for each programme. The second question of 'where we will provide opportunities' is considered through curriculum design, mappings to course learning objectives, and subsequent delivery of courses, to provide students opportunities to learn the knowledge, skills and values we have laid out in programme learning goals, programme learning objectives, and course learning objectives. The 'how will we know?' question means we need to assess to see whether the students

have learnt the desired learning objectives, and check whether there are gaps. The 'what will we change?' question 'closes the loop' and involves adjustments to programme elements or teaching methods in order to improve student learning outcomes where most needed. Periodically we would revise the mission, and/or programme learning goals, but for the most part, the cycle results in changes in one of the later stages.

The steps undertaken in answering each of the four questions are described in more detail below.

1. What do we want our graduates to know, do, and be?

The Faculty's Mission: to pursue and share knowledge of business, economics and management, to develop capability and provide our stakeholders with a global perspective, and our Vision: to be the scholarly hub driving New Zealand's capital development, provide a starting point.

From this we lay out learning goals for each of our programmes (degrees) such as our main undergraduate degree, the Bachelor of Commerce and Administration (BCA). For the BCA, the Learning Goals are:

Learning Goal 1: Critical and Creative Thinking

Our graduates will demonstrate application of critical and creative thinking skills to practical and theoretical problems

Learning Goal 2: Communication

Our graduates will be effective communicators

Learning Goal 3: Global and Multicultural Perspective

Our graduates will have a global and multicultural perspective

Learning Goal 4: Leadership

Our graduates will recognise, support and display leadership

Learning Goal 5: Major Attributes

Our graduates will develop specific knowledge and skills in at least one business, economics or public policy discipline area.

There is a clear and deliberate link between these Learning Goals and the Victoria Graduate Attributes that all Victoria graduates are expected to display, namely, Critical and Creative Thinking, and Communication and Leadership skills (Strategic Plan 2005-2014, Victoria University of Wellington). In addition, we have a Learning Goal relating to global and multi-cultural perspectives, in line with our mission, and we have also included 'major attributes' to cover discipline-specific learning goals.

The learning goals need further elaboration to aid their use for programme development and assessment. Accordingly, we list a set of operationally manageable learning objectives that are indicative of the graduate attribute we are aiming to develop. For example, for Learning Goal 1: Critical and Creative Thinking, the Programme-level Learning Objectives are:

Our graduates will be able to:

- a. analyse a complex situation which could be viewed from multiple perspectives;
- b. use/apply analytical techniques, models and/or frameworks appropriately in specific contexts;
- c. reflect critically on practical and theoretical issues;
- d. display creative thinking when faced with practical and/or theoretical problems.

A full list of learning objectives for all of the BCA learning goals is attached as Appendix 1.

for a particular major (just the compulsory courses are shown in the diagram, though we have mapped elective courses as well). Curriculum coverage is indicated by H, M, L (High, Medium, Low), while G indicates where the learning objective is assessed, contributing to a student's grade. The red shading is used to show where we decided to assess for AoL. Decisions on where to assess each learning objective for AoL purposes are made by considering where the learning objective is most likely to be demonstrated, along with workload implications for both staff and students. Assessments usually utilise existing coursework, with development/adaptation as needed. The curriculum maps quickly reveal where there are gaps or overlaps, and several Schools have used them to reorganise courses to more effectively and efficiently cover the various elements required.

3. How will we know if the students have acquired them?

The next step is to collect evidence on student achievement of the desired learning goals (graduate attributes). We need to assess student work to see whether there is evidence that they have learnt what we set out for them to learn (be able to do/ how they behave etc – i.e., a mix of knowledge, skills and attitudes/dispositions). This task is greatly aided by the curriculum maps, as they show where student work relating to specific learning objectives is already being assessed as part of determining students' grades. Such places should therefore be able to provide evidence on the levels of achievement of the particular learning objectives that contribute to the learning goal. Our aim is to ascertain the proportion of students who have demonstrated that they have achieved specific learning objectives, and hence learning goals.

For programme improvement purposes, it is sufficient to group students into just three categories, described by Martell (2009) as 'good enough, way good enough, and not good enough'. In our assessments, we use the terms Exemplary, Satisfactory and Unsatisfactory. We can then use this information to focus improvement efforts on the places (learning objectives, levels, or disciplines) where a significant proportion of students has not demonstrated an ability to reach the Satisfactory level.

The AoL system places a heavy emphasis on direct measures of learning, in two respects. Firstly, direct measures of learning are used in preference to indirect measures such as surveys, anecdotal evidence, focus groups, employer feedback, and student feedback. Indirect measures are considered by our accreditation bodies to be less reliable than direct measures, though they may be used to augment direct measurement, or where it is impossible to measure the learning objective directly (AACSB, 2007). Secondly, we cannot rely on grades, as these reflect overall performance on a combination of learning goals, and are unhelpful in assessing achievement on particular learning goals. Grades can also be influenced by relative performance of cohorts and often do not reflect absolute standards of performance. The relationship between student grades and the cohort's level of achievement on learning objectives is indicated in Figure 3. Assessors are asked to mark student work as usual to arrive at a grade (across a number of aspects being assessed), and then to assess it for AoL by collecting data on the student cohort's performance on one or more particular learning objective (indicated by the vertical arrows). The relevant rubrics are used for the AoL assessment, while the assessor's own marking scheme is used for marking. External assessors are used periodically to provide an independent assessment. The dovetailing of assessment for AoL with normal assessment with periodic audit provided something of a departure from most other AoL systems, and was arrived at as a result of the action research project using Theory of Constraints, described in Mabin (2009).

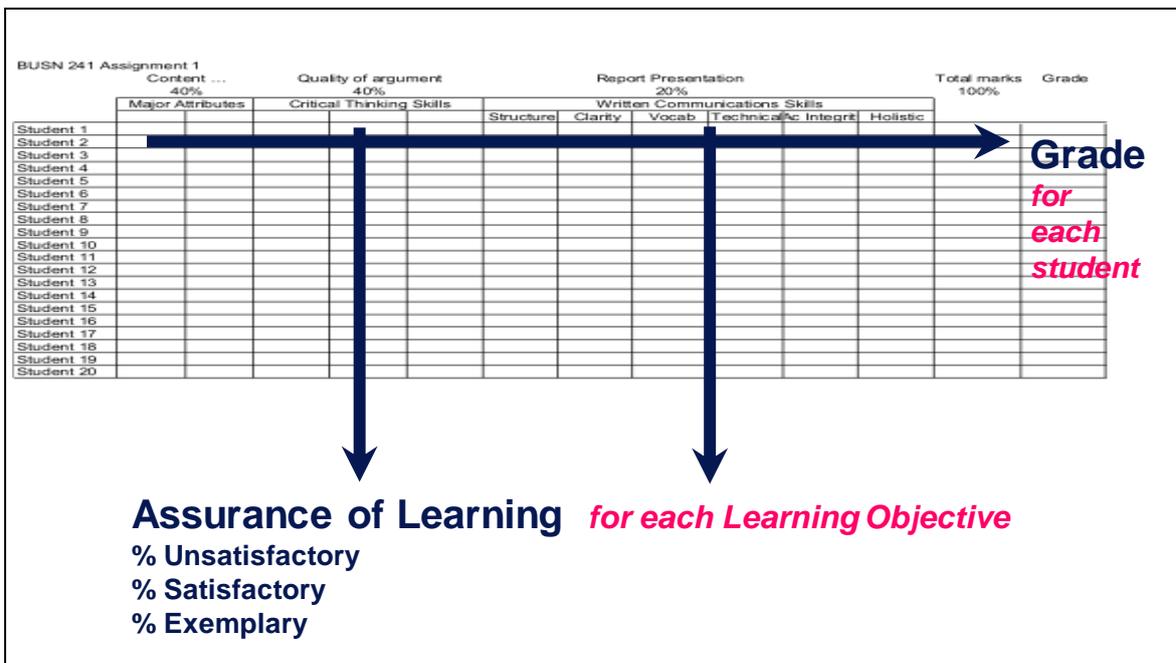


Figure 3: Assessing for grades vs. achievement of learning goals/objectives

Assessment plans were developed for all programmes for a set time period – the BCA, a 10-year plan, and for other degrees, a five-year plan. Each learning goal is being assessed at least twice in this period. For the BCA, we have already collected data over three years, with some on each learning goal, but with particular emphasis on Communication, and Critical and Creative Thinking. For each assessment, we take a representative sample of student work and assess it using an AoL rubric. Rubrics have been developed for communications skills (three rubrics: academic writing, business writing, and oral presentations using appropriate technology); for critical and creative thinking (four rubrics); and for leadership (rubrics for teamwork, and plan and lead a seminar or tutorial discussion). A sample rubric for one learning objective for Critical Thinking is shown in Figure 4. Rubrics have been developed for selected learning objectives relating to global/multi-cultural contexts, the role of ethics and public/private sector context and governance, and various Major attributes. Marksheets have been developed to assist with assessing research-related learning objectives as evidenced in masters theses. A selection of such rubrics and marksheets has been made available online (<http://www.victoria.ac.nz/fca/teaching/rubrics.aspx>).

Trait	Exemplary	Satisfactory	Unsatisfactory
Question assumptions	Identifies key assumptions of situation and disciplinary tools and discusses them in thorough and/or insightful fashion, validating, challenging, and/or revising as appropriate.	Key assumptions of disciplinary tools and situation are identified and clearly stated.	Assumptions are not identified, only tangentially discussed or implied rather than stated.
Draw conclusions	Draws thorough, appropriate conclusions demonstrating an ability to identify priority, significance and impact. Makes ethical judgements where appropriate	Draw appropriate, justifiable conclusions addressing major outcomes.	Conclusions do not follow from evidence and analysis, are far-fetched or trivial in scope.
Reflect	Comments in a structured and insightful way on: <ul style="list-style-type: none"> • Outcomes • Process • The initial situation • The disciplinary tools used • Their learning experience. 	Makes a substantive reflective statement addressing at least one of: <ul style="list-style-type: none"> • Outcomes • Process • The initial situation • The disciplinary tools used • Their learning experience. 	No attempt at reflection, or only superficial comments apparent.

**Figure 4: Sample Rubric
BCA Learning Objective LO1c: Reflect critically on practical and theoretical issues**

4. What will we (the Faculty) change if the students haven't learnt these things?

The results of each assessment are scrutinised by discipline specialists as well as the Associate Dean, in order to draw out any lessons for the future. Results are collated into a Faculty-wide summary report so that patterns can be identified for each learning goal, within and across the disciplines and levels. An extract from a summary report is shown in Figure 5. Discussions take place at the Faculty management level, in 'Teaching and Learning' committees, and back in the Schools. Suggestions are discussed and agreed and actions taken to 'close the loop,' with any common causes being addressed at the Faculty level.

Recommended actions have included redesign of rubrics, redesign of assignment question(s) or format so that students are led to demonstrate the learning objectives more explicitly, improvement to teaching methods or support provided, and better scaffolding or preparation in preceding courses. Faculty-wide actions have included writing skills development, and tutor training.

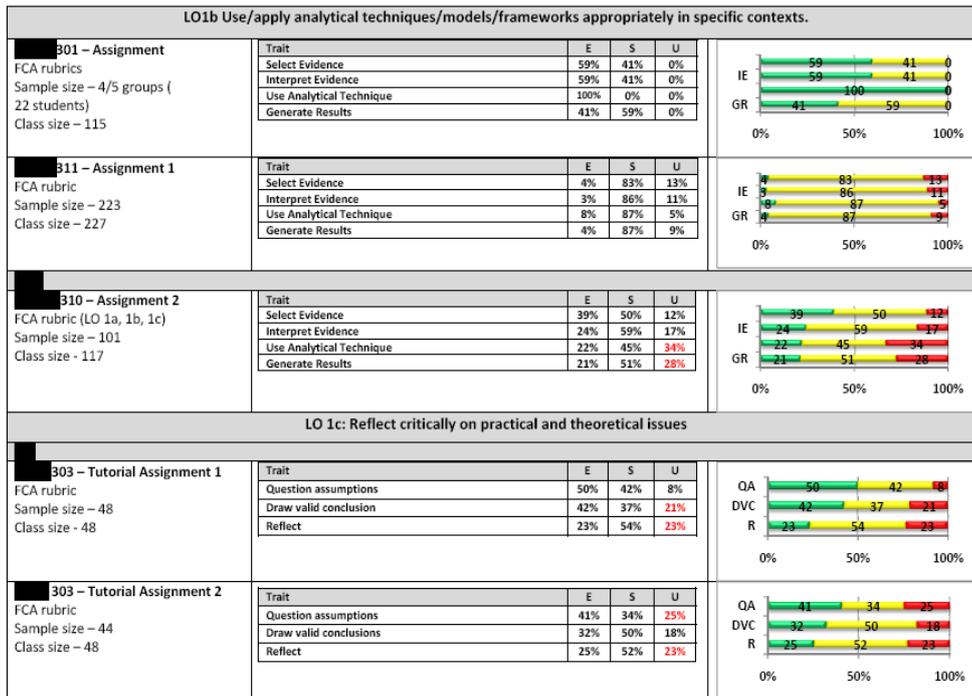


Figure 5: Extract from a Summary Report on AoL exercises

Discussion

The result of these developments is a feasible, working system, supporting continuous improvement via a cycle of planning, collecting data, learning and acting. Samples of student work are being assessed at various places throughout the degree to determine the levels of achievement of specific learning goals, allowing the Faculty to focus programme improvement efforts on the most needed areas. So far, efforts have been concentrated on one learning goal at a time, starting with communication (written and oral presentations), then critical and creative thinking, and now other learning goals, and have led to demonstrated local improvements in these areas, as discussed below.

Indicators of success include the achievement of accreditation from the AACSB, an achievement shared by only 0.3% of business schools worldwide. The peer review panel described the Faculty’s AoL as being ‘ahead of its peers’, and a top official wrote she was “very impressed how you took the ‘problem’ of establishing an AOL process and got it in place in what I would suggest was record time! We need other schools to realize how this can be done because some schools find it remarkably difficult to do this in any reasonable time. Also your data collection is efficient being real-time and in sync with the normal process of collecting grades” (E. Peacock, personal communication, March 9, 2011). This led to an invited presentation to business schools (Mabin, 2011). Internal recognition has been reflected in the voluntary uptake by two programmes that did not need to use AoL because they are officially ‘outside the scope’ for accreditation. The fact that these programmes chose to use AoL processes and found them useful, provides helpful endorsement to other colleagues within the Faculty.

The results of the processes introduced to date may be assessed in terms of the impacts and outcomes – both measured and observed – though as the system is being progressively developed, it is hard to evaluate year-on-year progress totally objectively.

Improvements in student learning outcomes through AoL have been achieved in each of the areas tackled so far. These include writing tasks, oral communication such as presentations, and critical and creative thinking. Firstly, there are many cases of

demonstrated improvement in writing quality following assessment and feedback using the Faculty's written communication rubric. Discussions on results are leading to sharing of best practice across the Faculty. For example, in several courses, staff have conducted successive writing exercises, with the rubric providing a framework for students and staff to discuss and clarify expectations and provide students with a way of deciding where they should focus their efforts. Improvements in student achievement have been noticeable. In one case, the writing in a second assignment was significantly better even though students were not told that their writing *per se* was being assessed separately in the second assignment. In another case, a mixture of student self-assessment using the rubric was then combined with staff feedback using the rubric, resulting in observable improvements in later written work. In oral presentations, students are performing well, and getting better with repeated practice, as with the writing skills exercises.

Critical thinking is also being progressively developed. Learning objectives were revised after failed attempts to devise rubrics for the original objectives. The resulting set of four revised learning objectives and their rubrics were trialled and then fine tuned, and are now being used widely with good results. The suite of rubrics has been found helpful in developing a shared understanding of what is meant by critical thinking. While this provides just one interpretation of critical thinking, the suite of rubrics developed has provided a common language and some benchmarks, which are proving to be useful for students and staff in laying out a framework for the achievement of this graduate attribute. Actions resulting from their use include improved assignment design, and more attention to important aspects like questioning assumptions, selecting, interpreting and using evidence, interpreting results and drawing conclusions.

The system uses criterion-based assessment using rubrics for each learning objective. The primary aim for AoL is to provide feedback to the teachers on how the student cohort is faring, in a formative rather than summative fashion. Poor performance may be addressed in future courses in a variety of ways, including changed teaching methods, or a more clearly worded assignment. However, students are also beneficiaries in gaining more clarity of expectations and feedback on their own performance.

Rubrics can form a useful basis of discussion for students and staff to clarify the task required and set high expectations, as well as providing useful feedback to students. Rust, Price, and O'Donovan (2003) recommend a lengthy discussion, but shorter discussions are also beneficial. Where the rubrics are marked up for every student, these marked rubrics can be returned to students, saving time for markers to provide directed personalised feedback. Where only a sample is marked using the rubrics, the teacher can use the statistics on the level of achievement of the student cohort to alert the class to common areas of concern and patterns of failure as well as strengths, and provide a stronger rationale for subsequent developments.

One early area of concern was written communication skills, and a Faculty plan of interventions was instigated. The Faculty provides academic writing skills tutorials alongside a 100-level core course, with specially trained tutors, and the performance of students in this course is assessed with AoL exercises. In addition, the effects of the writing skills course are being seen in later years. For example in a 200-level class where writing performance had been poor, we found that around 40% of students had not done the Faculty's 'writing skills course'. We also noted a correlation between the proportion of courses successfully completed and completion of the Faculty's 'writing skills course', which will be investigated further. Since 2010, all BCA students have been required to enrol in the 'writing skills course' in their first year of study, and the ongoing effects of this policy are being monitored.

The AoL process has also positively impacted on curriculum development, from our perspective and based on comments relayed by several School representatives. Early benefits of the process came from curriculum conversations which identified areas of overlap, repetition of material, and gaps in coverage, for a typical student in the major. As

to be expected, there has been a range of attitudes to this exercise. Some Schools in the Faculty recognised and grasped the opportunity for better coordination of effort, which assisted in the adjustment required for the changes in points value of courses, and in meeting concerns of staff and students regarding workload. These Schools are making more deliberate use of staircasing and scaffolding of student exercises within courses and between courses in the major. Several Schools have added capstone or compulsory courses at the 300-level to allow them to provide an integrative experience and build graduate attributes that may otherwise have been underdeveloped, as well as bringing the cohort together. There is still a long way to go before we have a curriculum that builds all learning goals, in an interlocked fashion, in all majors, but at least we have started the journey.

The Schools responding favourably to the opportunity to review their curricula have also taken on the AoL tasks with ease, and have readily incorporated the AoL tasks into their normal assessment regimes in quite remarkable fashion. At the other extreme, some staff have regarded the curriculum mapping merely as a chore, according it perfunctory compliance, and gaining little from the exercise. However, all Schools are undertaking assessments, and are starting to gain more familiarity and confidence with the process.

The development, usage and acceptance of rubrics over time have been interesting to observe. We have developed a growing suite of Faculty-agreed rubrics for generic skills, as well as tailored rubrics for major-specific attributes. Rubrics for communication skills, and critical thinking are being widely used, and for teamwork and creative thinking, and a marksheet for thesis research is in use, while rubrics for global/multicultural perspectives are being developed.

Rubrics have clarified expectations for staff as well as students. Some teachers have used these rubrics for AoL after normal marking using their own marking schemes, as was intended. Others have integrated rubrics into their own marking schemes to good effect. However, some have done so rather too well, adding weights, and using them for grading in place of their own marking schemes. Too tight a connection between rubrics for AoL and their use for grading is not advised (Sadler, 2009) and indeed has led to concerns being raised by students. Balanced/appropriate use of rubrics is therefore being urged, to avoid overuse or misuse of rubrics.

We have been careful with our rubrics to avoid some of the known drawbacks seen in other rubrics. In particular there are several key respects in which our rubrics differ from those used in New Zealand's National Certificate in Educational Achievement (NCEA). For example, our rubrics include only two 'passing' categories, whereas NCEA makes finer distinctions with three. NCEA rubrics are used directly for grading, whereas ours are used primarily for gauging the level of achievement of the cohort, based on a sample of students. NCEA rubrics do not allow the assessor any discretion to compensate for poor performance on one trait by exemplary performance on another trait. We have added a holistic trait to many of the rubrics to allow the assessor to make an overall judgement, bearing in mind the balancing out of inadequate and excellent performance, as well as any other factors in determining effectiveness as a whole. This holistic judgement also acknowledges that the rubric traits may not completely capture every aspect of the learning objective being assessed.

The AoL system has been developed over the last three years, and is now used for conducting assessments in the BCA and other programmes at Masters and Bachelor levels, including Masters research by thesis. We have made many improvements to the system itself over the three years – streamlining the process, dovetailing AoL assessments with normal marking, and providing materials and resources on the Faculty's newly launched Teaching Matters website (see <http://www.victoria.ac.nz/fca/teaching/aolresources.aspx>).

As with any new initiative, we have made many mistakes. Improvements to the system are just as much a part of the process as continuous improvement of student learning outcomes.

- Some assessments used group reports to assess learning objectives for which group reports are inappropriate. While group reports may be a good indicator of the effectiveness of group work, they do not provide a reliable indicator for assessing the proportions of students who have achieved learning objectives such as writing skills or critical thinking. A satisfactory score merely indicates that at least one student in the group achieved this level; valid inferences cannot be made about the cohort as a whole. This has led to some interesting observations though, such as when the same class's group projects scored worse than individual reports, which was attributed to a lack of coordination and consistency in writing and presentation style.
- In other cases, the rubric was not tuned or the assessment was not suitable, resulting in too many Exemplary scores, or too many Unsatisfactory scores. Such results may be prevented by a pilot, or by better moderation. In other cases, it may be necessary to revise the assignment so students are able to demonstrate their abilities well.
- Some assessors found it hard to decide on a category, and assigned a borderline score. AoL is not grading! For AoL, we are really only interested in the proportions of the cohort in each category, which only needs to be approximately right.
- Some teachers did not take copies of student work before handing it back to students, underlining the need to have systems in place to catch it in time.

Improvements for the future are to encourage staff to use the results with their classes, and we are also noting course pass rates alongside the percentages of Exemplary, Satisfactory and Unsatisfactory, as an aid to interpreting the percentage of students falling in the Unsatisfactory category for each AoL exercise. Assessment results are being collated and will be reported in future papers, and results from Victoria are being compared with results from other institutions, particularly other Australasian business schools.

Conclusions and Recommendations

We have developed a system that is working effectively, providing us with information on the levels of achievement of the various learning goals, and allowing the Faculty to focus on programme improvements. So far, efforts have been concentrated on communication (written and oral presentations) and critical and creative thinking, and have led to demonstrated local improvements in these areas. We are trialling activities to develop global/multi-cultural perspectives, taking a coordinated approach to addressing gaps in the curriculum. In addition to providing feedback to teachers on how the student cohort is faring, students are also benefitting from greater clarity of expectations and feedback on their performance.

The proactive approach to assuring student learning also provides greater accountability to external stakeholders, including students, advisory boards, accreditation agencies, funding agencies and quality monitoring agencies such as TEQSA, NZQA and TEC.

Our next steps include: 1. developing more clarity and consistency of interpretation of the rubric standards, using advisory panels to advise on standards, and improved training of tutors to provide better consistency of assessments, an issue that is also being researched in Australia and USA (Baker, 2011; Freeman, 2011); 2. a more streamlined process for collating and analysing data, and compiling and disseminating results; 3. revisiting curriculum maps and using constructive alignment in curriculum planning on a regular basis and with greater emphasis on 'interlock'; and 4. continuing to build the understanding, acceptance and uptake of AoL. These developments will continue to form

part of the broader action research project, with effects being measured through AoL assessments, wherever appropriate, with data being collected and collated for future publication. Ongoing developments at Victoria will inform and be informed by comparable work at other business schools.

While such developments indicate that the system is still work in progress, the AoL processes developed for business schools are proving to be very useful in ways that could enhance other programmes, and deserve wider consideration. In particular, universities with graduate attributes may find that similar processes would enable better-informed curriculum decisions as well as claims about student achievement of graduate attributes. The processes described provide a way of ensuring that graduate attributes are developed and assessed in each student's programme, so that universities can make evidence-based claims about their students' capabilities. When carefully designed to utilise existing content and coursework, the overheads of a superimposed system can be lessened. Furthermore, it empowers institutions to develop systems and standards to suit their own mission, rather than being forced into a one-size-fits-all standards model.

Acknowledgements

The authors wish to acknowledge the invaluable contributions of their many colleagues in developing and implementing the system described.

References

- AACSB International (The Association to Advance Collegiate Schools of Business). (2007). *AACSB assurance of learning standards*. Retrieved from <http://www.aacsb.edu/accreditation/papers/AOLPaper-final-11-20-07.pdf>
- Baker, E. L. (2011, September). *Assessment in a changing world*. Keynote Address, Symposium on Assessment and Learner Outcomes, Victoria University of Wellington.
- Banta, T. (2008, November). *Profiles of good practice in assessing student learning outcomes*. Keynote address. ATE and HELT Conference, Victoria University of Wellington.
- Barnett, R., & Coate, K. (2005). *Engaging the curriculum in higher education*. Maidenhead: Open University Press.
- Barrie, S. C. (2004). A research-based approach to generic graduate attributes policy. *Higher Education Research & Development*, 23(3), 261-275.
- Biggs, J. B. (1999). *Teaching for quality learning at university*. Milton Keynes: Open University Press.
- Bowden, J., Hart, G., King, B., Trigwell, K., & Watts, O. (2000). *Generic capabilities of ATN university graduates*. Retrieved from <http://www.ct.uts.edu.au/TheProject.htm#one>
- Clanchy, J., & Ballard, B. (1995). Generic skills in the context of higher education. *Higher Education Research and Development*, 14, 155-166.
- Coats, M. (2000). *Compliance or creativity? Using learning outcomes to enhance student autonomy*. Conference on Academic Development: Challenges and changes, Rhodes University, South Africa. Retrieved from <http://www.open.ac.uk/cobe/resources/RhodesSA.pdf>
- Connor, P. E. (1997). Total Quality Management: A selective commentary on its human dimensions with specific reference to its down side. *Public Administration Review*, 57(6), 501-509.
- Crebert, G. (2002). *Institutional research into generic skills and graduate attributes: Constraints and dilemmas*. Paper presented at the International Lifelong Learning Conference, Yeppoon, Queensland.

- Davies, J. (2010). Management theory: A systems perspective on understanding management practice and management behavior. *International Journal of Strategic Decision Sciences*, 1(3), 33-48.
- De Jager, H. J., & Nieuwenhuis, F. J. (2005). Linkages between Total Quality Management and the outcomes-based approach in an education environment. *Quality in Higher Education*, 11(3), 251-260.
- De La Harpe, B., Radloff, A., & Wyber, J. (2000). Quality and generic (professional) skills. *Quality in Higher Education*, 6, 231-243.
- Deming, W. E. (1982). *Out of the crisis*. Cambridge, MA: MIT Center for Advanced Engineering Study.
- Dillon, C., Reuben, C., Coats, M., & Hodgkinson, L. (2007). Learning outcomes and their assessment: Putting Open University pedagogical practices under the microscope. In S. Frankland (Ed.), *Enhancing teaching and learning through assessment: Deriving an appropriate model* (pp. 280-289). Dordrecht, The Netherlands: Springer.
- Freeman, M. (2011, September). *Disciplines setting and assessing academic standards*. Victoria University of Wellington UTDC Seminar.
- Globalization of Management Education Task Force. (2011). *Recommendations to AACSB International from the Globalization of Management Education Task Force*. Retrieved from <http://www.aacsb.edu/publications/researchreports/>
- Goldratt, E. M. (1994). *It's not luck*. Great Barrington, MA: North River Press.
- Higher Education Council. (1992). *Achieving quality*. Canberra: Australian Government Publishing Service.
- Hoecht, A. (2006). Quality assurance in UK higher education: Issues of trust, control, professional autonomy and accountability. *Higher Education*, 51, 541-563.
- Houston, D. (2007). TQM and higher education: A critical systems perspective on fitness for purpose. *Quality in Higher Education*, 13(1), 3-17.
- Koch, J. V. (2003). TQM: Why is its impact in higher education so small? *The TQM Magazine* 15(5), 325-333.
- Mabin, V. J. (2009, December). Eating the elephant! Applying theory of constraints in assurance of learning. *Proceedings of the 44th Annual Conference of the Operational Research Society of NZ* (pp. 282-291). Christchurch.
- Mabin, V. J. (2011, May). *Developing effective assurance of learning goals, measures and processes*. AACSB Asia Pacific Regional Conference, Singapore.
- Mabin, V. J., Forgeson, S., & Green, L. (2001). Harnessing resistance: Using the theory of constraints to assist change management. *Journal of European Industrial Training*, 25(2/3/4), 168-191.
- Martell, K. (2009, March). Assessment seminar, AACSB International, Paris.
- NCIHE. (1997). *Higher education in the learning society*. (Report of the National Committee of Inquiry into Higher Education, 2 Vols.). London: HMSO.
- Oliver, B. (2011). Assuring graduate outcomes. Australian Learning and Teaching Council. Retrieved August 14, 2011, from <http://www.altc.edu.au/resource-assuring-graduate-outcomes-curtin-2011>
- Rust, C., Price, M., & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education* 28(2), 147-164.
- Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, 34(2), 159-179.
- Sumsion, J., & Goodfellow, J. (2004). Identifying generic skills through curriculum mapping: A critical evaluation. *Higher Education Research & Development*, 23(3), 329-346.
- The National GAP (Graduate Attributes Project), Institute for Teaching & Learning, University of Sydney. Retrieved from <http://www.itl.usyd.edu.au/projects/nationalgap/resources/gamap/map.cfm?ga=5> accessed 23/08/2011
- Treleaven, L., & Voola, R. (2008). Integrating the development of graduate attributes through constructive alignment. *Journal of Marketing Education*, 30(2), 160-173.

Appendix 1. BCA Learning Goals and Learning Objectives

Learning Goal 1: Our graduates will demonstrate application of critical and creative thinking skills to practical and theoretical problems (“Critical and Creative Thinking”)

Learning objectives. They will be able to:

- a) Analyse a complex situation which could be viewed from multiple perspectives.
- b) Use/apply analytical techniques/models/frameworks appropriately in specific contexts.
- c) Reflect critically on practical and theoretical issues.
- d) Display creative thinking when faced with practical and/or theoretical problems.

Learning Goal 2: Our graduates will be effective communicators (“Communication”)

Learning objectives. They will bring creative skills to the research, analysis, planning, and writing stages of academic essays and other documents, and will be able to present them clearly and effectively to an audience. That will involve being able to:

- a) research, plan, and produce written assignments meeting academic standards;
- b) apply advanced written communication skills in a private or public sector ‘business’ context;
- c) deliver a professional quality presentation accompanied by appropriate technology;
- d) demonstrate oral communication skills by their participation in small group learning environments such as tutorials and workshops.

Learning Goal 3: Our graduates will have a global and multicultural perspective (“Global and Multicultural Perspective”)

Learning objectives. They will be able to:

- a) define key components of countries’ business environments and give examples of how they differ across countries;
- b) adapt a domestic strategy to a foreign operational setting and/or adapt a foreign strategy to a domestic setting;
- c) assess the way in which legislation and government policy influence the business environment in national and global contexts;
- d) diagnose cross-cultural communication issues in a case setting and propose appropriate solutions;
- e) demonstrate an awareness of cultural differences and the skills needed to work effectively in multi-cultural environments within New Zealand and internationally.

Learning Goal 4: Our graduates will recognise, support and display leadership (“Leadership”)

Learning objectives. They will be able to:

- a) plan and lead a seminar or tutorial discussion;
- b) demonstrate an understanding of the relevance of ethics to the public and private sectors, and of the role of ethics in public and private governance;
- c) work constructively in groups.

Learning Goal 5: Our graduates will develop specific knowledge and skills in at least one business, economics or public policy discipline area (“Major attributes”)

Learning objectives as specified for majors, that is, each major in the BCA specifies several discipline-specific attributes that each of its graduates is to possess.

Schools Out – Gone to the Beach: How Some Year 11 New Zealand High School Students Perceive Study Leave

Peter Sanders
Victoria University of Wellington

Email: peter.sanders@vuw.ac.nz

Abstract

The under-achievement of some students in New Zealand secondary schools is of concern to parents, educators, academics, and policy-makers because of evidence that Pasifika students are disproportionately likely to have lower achievement outcomes than students from other ethnic groups. One strategy used by schools to address under-achievement involves study skills programmes offered to students during the school year. This paper reports on responses from Year 11 Pasifika students who took part in study skills programmes. Students were interviewed about strategies they used to prepare for their Level One National Certificate of Educational Achievement (NCEA) external examinations to determine what factors were seen by students as either supports or barriers to their study efforts. Findings indicated that students already at risk of under-achievement were further disadvantaged by not understanding policies and practices that had a direct impact on their performance. Some students believed they could not ask for legitimate assistance during assessments or did not understand the derived grade process.

Students also found it difficult to resist spending the examination study period 'at the beach' with their friends rather than focused on study when freed from school to go on study leave for as much as four weeks of term time. These findings indicate that more can be done by schools to improve learner outcomes for under-achieving students both in communicating examination policies and ensuring that study time was spent at school rather than assuming they had the support needed to be sent away from school to prepare for examinations. This paper concludes with recommendations to schools about how to support under-achieving students more effectively during a period of study time.

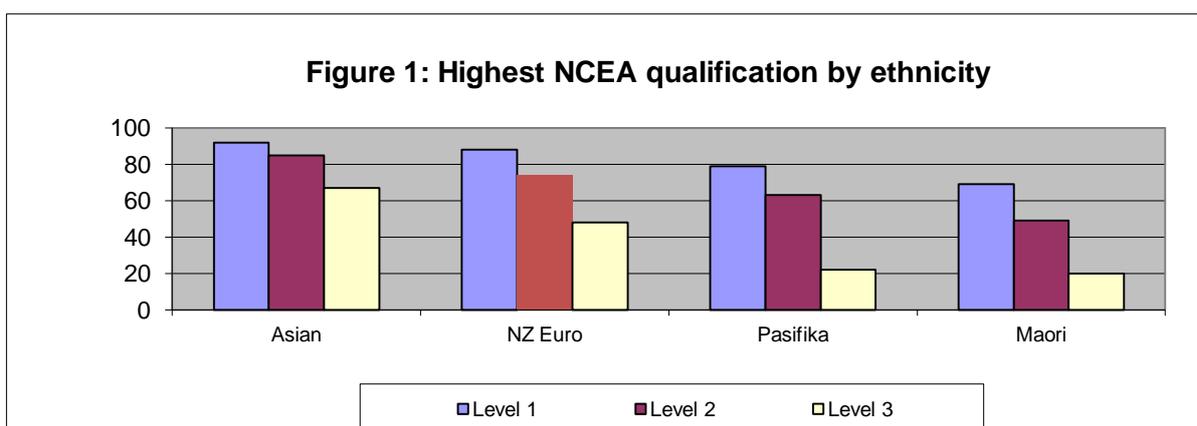
Introduction

The under-achievement of some students in New Zealand secondary schools is of concern to educators, academics, and policy-makers in the Ministry of Education. Of particular concern are ethnic patterns that show that Pasifika students have lower achievement outcomes than students from other ethnic groups. Biddulph, Biddulph, and Biddulph (2003) analysed all the major national and international assessment studies used to monitor achievement in New Zealand schools, by ethnicity. The five major national and international assessment studies used to monitor achievement in New Zealand schools were the Third International Mathematics and Science Study for Year 5 students; the National Education Monitoring Project (NEMP) for Years 4 and 8 students; the School Entry Assessment for Year 1 students; the International Association for the Evaluation of Educational Achievement on reading achievement for students at Year 5; and the Programme for International Student Assessment for Year 11 students. Biddulph et al. concluded that in every one of these studies the achievement of Māori and Pasifika students was significantly lower than that of European and Asian students.

Pasifika students show lower levels of academic achievement compared to students from other ethnic groups at a relatively early age. The NEMP (Ministry of Education, 2008) Reading and Speaking website noted that "Year 4 Pasifika students scored moderately lower than Pākehā students in both reading and speaking, a disparity which has decreased a little over the past eight years, but the corresponding differences for year 8 students are quite large and not decreasing." Phillips, McNaughton, and McDonald (2001) also reported a gap in literacy after four years at school, evidencing significant differences

in achievement between Māori and Pasifika students, and other students. This ethnic achievement pattern continues as students progress to secondary education, with Māori and Pasifika students leaving school with significantly lower qualifications than Asian and European students.

The Education Counts (2009) website reported there had been improvement in achievement for Māori and Pasifika students between 2002 and 2008 and concluded that disparities between ethnic groups were reducing. However, further analysis of Education Counts (2009) data shows that although the achievement differences between ethnic groups may be decreasing at Level 1 of the National Certificate of Educational Achievement (NCEA), fewer Pasifika and Māori students left secondary school in 2008 with Level 3 NCEA, compared to Asian and NZ European students. Currently just one in five Pasifika students leaves school with NCEA Level 3.



Source: Education Counts (2009)

Robinson and Timperley (2004) noted that there was pressure on the Ministry of Education to improve the achievement of Pasifika students and that this was an urgent issue. One measure adopted by schools to address under-achievement is the provision of a study programme. While there is little empirical evidence in the New Zealand context on the effectiveness of such programmes, Gorinski (2005) reported that the provision of a study centre at some Pacific Island School-Parent-Community Liaison Programme schools “had been an effective mechanism for supporting student achievement” (p. 18).

Literature Review and Theoretical Perspectives

This section reviews the literature about motivation achievement theory, showing how the relationship construct can be viewed as an overarching concept that links motivation theories. This section also reviews literature about the importance of relationships to general student achievement at school. Research about motivation and achievement in a New Zealand context is also reviewed as well as studies about the effects of motivation intervention programmes in Australian high schools and studies that give advice about the design of successful study skills programmes. Emphasis is placed on literature which focuses on the achievement of minority group students because this study investigates the effects of motivation interventions as a new way to address the under-achievement of some Pasifika students in New Zealand secondary schools.

The theoretical framework for this study rests on six key motivation theories. These are the attribution, expectancy-value, goal, self-determination, self-efficacy, and self-worth motivation theories (Martin & Dowson, 2009). Martin (2001) believes that “an understanding of each of these individual theories will assist educators to improve their understanding of why, how, and the way, students are motivated to achieve” (p. 1).

Attribution motivation theory explains 'why' an event has occurred (Weiner, 1985) and indicates that students frequently make four common attributions for their success, or lack of success, in assessments: luck, task difficulty, ability, or effort (Martin & Dowson, 2009). Attributions are considered to fit into a three-dimensional model composed of locus, stability, and control (Weiner, 1985). Locus refers to the attribution being internally or externally located in a student's mind. Ability and effort are considered internally positioned, whereas task difficulty and luck are externally located. Stability refers to how open the attribution is to educational amendment, and control refers to a student's belief about how their preparation for a particular assessment affected the subsequent outcome (Martin & Dowson, 2009; Meyer, Weir, McClure, Walkey, & McKenzie, 2009; Ng, McClure, Walkey, & Hunt, 1995; Weiner, 1985). For example, if a student attributes a poor test result to lack of effort, they may believe that with more effort they could do better in another test situation. Effort is considered easy to change, less demotivating, perhaps even motivating. Conversely, if the student attributes the poor test result to their own low level of ability, this attribution may be considered less easy to change and more demotivating.

The concept of relationships may be connected to attribution motivation theory by examining teacher feedback to students. Following a classroom assessment situation, a teacher may give positive feedback such as "Well done, you must have worked hard." The student may then attribute their grade to their effort and develop increased motivation to achieve. Conversely, negative feedback, such as "Did you even study for the test?" could cause negative attributions about results, and subsequent student de-motivation to achieve (Martin & Dowson, 2009, p. 334).

Expectancy-value motivation theory explains that motivation is the product of motive, multiplied by expectancy of success, multiplied by incentive to succeed (Atkinson, 1957). Martin and Dowson (2009) explained that students who have higher motivation, if they value the task and have a high expectation they will succeed. Again the relationship construct can be associated with expectancy-value motivation theory with research showing that teachers who expected less of particular students gave them correspondingly less demanding work (Wigfield, Galper, Denton, & Seefeldt, 1999). Wigfield et al. also reported that teachers could accurately predict the achievement level a student would attain in advance and concluded that this proved teacher expectation was a valid predictor of student achievement.

However, Rubie-Davies, Hattie, and Hamilton (2003) had problems with the teacher expectation concept because they showed some teachers predicted lower achievement for Māori students because the students were Māori. This was despite those same teachers using pre-test data that showed Māori students had the same achievement levels as students from other ethnic groups. They concluded that teacher expectation was actually a self-fulfilling prophecy because subsequent Māori students' achievement was worse than other ethnic groups.

Goal motivation theory relates achievement to the meanings and goals students attach to assessments and is useful in the educational context because it 'locates' where a student is on an 'approach-avoidance' continuum. For example, a student who willingly chooses to participate in a classroom activity would be considered more positively motivated than a student who avoids taking an assessment because they think they will fail. Relationships are associated with goal motivation theory because a good teacher-student relationship can assist a sense of 'belonging' in the classroom, which is fundamental to student achievement (Martin & Dowson, 2009).

Self-determination motivation theory is based on the concept that a student's psychological needs of relatedness, competence and autonomy must be met before achievement can occur. The relationship concept is relevant to this motivation theory because a student's perception of the teacher-student relationship will affect their self-determination, motivation, and subsequent achievement (Martin & Dowson, 2009).

Self-efficacy motivation theory is believed to account for some of the observed variance in students' academic achievement whereas Bandura (2007) showed that students with high self-efficacy could find alternative solutions more easily, made more effort, and were more persistent at tasks compared to students with lower levels of self-efficacy who tended to focus on their deficiencies (Martin, 2001). It is believed that self-efficacy can be enhanced through positive teacher-student relationships because self-efficacy is usually improved through positive relationships with others in a social setting (Martin & Dowson, 2009).

Self-worth motivation theory is important to a student's sense of their ability. Self-worth improves when effort in the classroom results in the student gaining confidence in their ability to succeed. Conversely, there are negative consequences for self-worth if a student does try hard but is not successful as this can lead to the student concluding they have low ability (Covington & Omelitch, 1980).

There is common ground between self-worth theory and attribution theory because students can view ability as being fixed or open to change. If a student does not believe that ability equates to changeable skills, they may begin to believe they are not 'smart' and sink into "maladaptive helpless orientation" (Meyer, McClure, Walkey, McKenzie, & Weir, 2006, p. 9; Dweck, 1992). Martin (2007) viewed motivation orientations as being adaptive or maladaptive. Again the concept of relationship can assist understanding of self-worth motivation theory where Martin and Dowson (2009) reported that good parent-child relationships were important because a child models their behaviour on the way they see their parents dealing with their own self-worth issues. Wentzel (1999) also associated self-worth motivation orientation with relationships and showed that teacher approval correlated with a student's sense of self-worth.

Relationships and Achievement

In addition to the association between relationships and motivation, there is a considerable field of literature that reports the importance of relationships to general student achievement at school. The literature shows that both teacher-student and parent-child relationships are important, although teacher-student relationships have greater impact on achievement than parent-child relationships. Martin, Marsh, McInerney, Green and Dowson (2007) analysed the role of teacher-student and parent-child relationships in students' academic motivation and engagement, self-concept and general esteem. They found that while both types of relationships were important to these issues, teacher-student relationships were most strongly associated with academic concepts.

The effect a teacher has on student achievement has been quantified. Hattie (2003) reported that six factors were related to observed variation in student achievement: curricula, policy, school climate, the teacher, teaching strategies, and the home. Hattie calculated the effect size of each factor on student achievement and concluded that students themselves account for 50% of the observed variance in achievement, teachers contribute 30%, and the home affects just 5–10% of the variation. Hattie concluded that apart from the students themselves, the teacher makes the greatest difference in variance in student achievement.

These results may also be associated with findings from Richer, Godfrey, Partington, Harslett, and Harrison (1998) who reported that a high percentage of Aboriginal students experienced relationship problems with their teachers. They surveyed 473 Aboriginal students about student perceptions of teacher attitudes and reported that 37% disagreed-strongly or disagreed with the statement "my teacher cares what happens to me", and 58% disagreed-strongly or disagreed with the statement "I like the teacher." While no data are produced in this report about the attitudes of corresponding non-Aboriginal students, it could be concluded that a significant percentage of Aboriginal students in this study had a relationship issue with their teachers. Low academic outcomes of Aboriginal students may be associated with these reported poor relationships with their teachers. Martin (2006) believes that relationships are important to Aboriginal student achievement on three levels: between the teacher and student,

between the student and the school, and between the student and pedagogy. Martin also noted that teachers need to be warm, have positive expectations, get to know their students and respect them as individuals to allow a good relationship to develop and avoid a poor relationship that could impede achievement. Martin's (2006) study has commonalities with the *Te Kotahitanga* project, as reported by Bishop, Berryman, Cavanagh, and Teddy (2010), which showed that relationships were important to the achievement of Māori students in New Zealand secondary schools. Alton-Lee (2003) concurred, stating that "The quality of classroom relations and interactions within schools has more to do with the creation of educational disparities than the decile ranking of the schools" (p. 7).

Motivation and Achievement in a New Zealand Context

Meyer et al. (2006, 2007, 2009) researched motivation and achievement in New Zealand secondary schools in the context of the NCEA. Meyer et al. (2006) found that negative motivation orientations were associated with lower achievement, reporting that "the strongest predictors of high academic achievement and higher grades were a high motivation orientation towards *Doing My Best* and a low motivation orientation towards *Doing Just Enough*." They pointed out that students who "do just enough" are at risk of failing from lack of effort and not ability (p. 2).

Meyer et al. (2006) also found that ethnicity was associated with achievement reporting that "Pasifika students gained fewer achievement standard credits than all other ethnic groups" (p. 39). The quality of grades was also an issue as "Māori and Pasifika students gained significantly fewer achievement standards with Merit than Asian and European students" and "Māori and Pasifika students gained fewer achievement standard credits with Excellence than all other ethnic groups" (p. 40).

Meyer et al. (2009) discussed the importance of teacher-student relationships and motivation noting there was "a significant relationship between the motivation dimensions and relationships with peers and teachers" (p. 2). They also noted that "students high on *Doing Just Enough* reported that their teachers did not take a personal interest in their achievement, whereas students high on *Doing My Best* reported that teachers showed interest in them and in their work" (p. 2).

Earlier research by Meyer et al. (2007) included a follow-up study of the two motivation orientations they labelled as *Doing My Best* and *Doing Just Enough*. They reported that "*Doing My Best* and *Doing Just Enough* were the strongest predictors of subsequent school achievement" (p. 2). Meyer et al. developed a screening tool and reported that their findings supported the "predictive validity and utility of the screening tool" (p. 2).

Meyer et al. (2007) also reported on associations between motivation orientations and attribution theory noting that "students who attributed their best work to internal factors of ability and effort showed the most positive achievement pattern overall and were most likely to report the *Doing My Best* orientation" (p. 3). Conversely, they reported that "students reporting a *Doing Just Enough* orientation were more likely to attribute best marks to luck and worst marks to a lack of ability. They were more likely to attain credits with Achieved rather than Merit or Excellence" (p. 3). Meyer et al. concluded that students "will have little motivation to exert more effort in future tasks and opportunities unless strategies are identified to change these motivations and attributions" (p. 3).

Meyer et al.'s (2007) findings correlate with those of Jones (1991) who investigated the question "Where does learning ability come from?" with female Pasifika students at a secondary school in New Zealand. Pasifika students attributed their lack of academic success almost entirely to a lack of "brains" (ability), rather than to effort. Alton-Lee (2003, p. 34), citing Nuthall (1999), considered that "Ability appears to be the consequence, not the cause, of differences in what students learn from their classroom experiences" (p. 213). Nuthall seems to be indicating that cultural mismatch may be associated with Pasifika students' attribution of ability.

Motivation Interventions

Meyer et al. (2009) reported that students' motivation attributions for their performance can be modified and Martin (2005, 2008) reported on two intervention programmes that aimed to change motivation orientations. Martin (2002) conceptualised achievement motivation by amalgamating the key theories of academic motivation into a multi-dimensional model called "The Student Motivation Wheel" which incorporated six major motivation achievement theories into an intervention programme. Martin (2005) then investigated the impact of this intervention programme on participants' academic motivation and found that the levels of some adaptive motivation constructs were higher, both immediately after the intervention, and 6–8 weeks later. He emphasised that the results of this study were significant in showing that even a brief but well-targeted intervention could yield significant positive changes in motivation. Martin (2008), reporting on another intervention, noted an improvement in targeted aspects of motivation constructs as evidenced by significant effect sizes ranging from 0.38–0.64 in the treatment group, in comparison to the control group.

Martin (2008) suggested that further research to test associations between changes in motivation and changes in achievement would be useful. Meyer et al. (2006) reported that intervention programmes, which impacted on student motivational orientations, could affect long-term student learning outcomes. Similarly, Meyer et al. (2009) noted that "An approach to intervention that highlighted intrapersonal motivation orientations could add significant value to the more traditional approach of academic remediation only; Martin's work in Australia provides an excellent example of this (Martin, 2008)" (p. 104).

Study Skills Intervention Programmes

Hattie, Biggs, and Purdie (1996) investigated the effectiveness of study skills programmes and concluded that "the best results came when (study skills) strategy training was used metacognitively, with appropriate motivational and contextual support." (p. 99). These points provide a theoretical base for designing effective study skills intervention programmes.

Contextual support refers to how closely the study programme is associated with the subject matter to be learned. Hattie et al. (1996) reported that "if strategy training is carried out in a metacognitive, self-regulative context, in connection with specific content rather than generalized skills, and if such training is supported by the teaching context itself, positive results are much more likely" (p. 101). This means schools should aim to deliver study skills programmes within the context of a particular programme.

Hattie et al. (1996) explained that "meta-cognitive interventions work on self-management skills such as planning; monitoring; and where, when and how to use tactics and strategies" (p. 189). Again this aspect of a study programme will be familiar to teachers and more detail is outlined by Martin (2009) who described how to teach planning skills. The third and final design aspect of a study skills programme is related to motivation. Hattie et al. (1996) explained that "affective interventions are those that focus on such non-cognitive aspects of learning as motivation and self-concept" (p. 100). They defined attributions for success and failure as affective. This part of a study skills programme would require careful design as the concepts covered here may be outside the scope of the regular teacher's skills. However, suitable professional development could lead to all classroom teachers being able to deliver the motivation aspect of a successful study skills programme. The importance of motivation to a study skills programme should not be underestimated as Hattie (2009) noted that "it can take less effort by a teacher to demotivate students compared to the often greater effort required to motivate them – to turn students on to learning" (p. 48).

Ethics and Cultural Advice

This research followed the *Pasifika Education Research Guidelines: Final Report* (Anae, Coxon, Mara, Wendt-Samu, & Finau, 2001) to ensure appropriate culturally sensitive behaviour was followed when designing and undertaking this study. Dickie (2008) explained that this should include “consultations, using translations, face-to-face meetings and the need to build rapport with participants who are being interviewed” (p. 81). Dickie made it clear that this means the data are not so much being collected but are instead co-authored; this is particularly important when the researcher is papalagi, as is the case here. The researcher consulted with three experienced Samoan educators who had agreed to be cultural advisors to guide the interview questions, protocols for the investigation, data collection, and the process of gaining consent for student participation and on-going consultation about organising the students in the project, as recommended by Dickie.

Method

A two-phase, sequential, explanatory, mixed-method research design (Creswell, 2009) was used to compare the effects of a Motivation-enhanced Study support programme (MS) with the effects of a Traditional Study support programme (TS). In the first phase, quantitative research questions addressed the relationship between achievement outcomes of participants in the MS programmes to those of participants in the TS programme. Information from this first phase was explored further and was used to inform the second qualitative phase. In this second stage, qualitative interviews were undertaken to listen to the voices of these students to improve the researcher’s understanding of achievement-related outcomes and to hear students’ perceptions of how the interventions did or did not support their academic performance. In this way, findings from the quantitative phase were enhanced by gaining the participants’ views about the two study support programmes. Most of the data reported and discussed in this paper come from interviews with students. Bishop et al. (2010) and Cook-Sather (2002) reported that students can feel empowered if teachers listen to and learn from students. Improved educational practice occurs when teachers listen to, and learn from, students as teachers can begin to see the world from the perspective of students. Cook-Sather reported that “authorising student perspectives is essential because of the various ways that it can improve educational practice, re-inform existing conversations about educational reform, and point to the discussions and reform effects yet to be undertaken” (p. 3).

Participants

Students were recruited from two secondary schools in a similar geographical area with School A categorised as a decile¹ 6 school and School B, decile 1. There were 57 participants, all Year 11 students, aged around 15 years, with 28 female students and 29 male students. There were 8 Māori, 28 Pacific, 18 NZ European and 3 NZ Asian/Asian students in this study.

Random Assignment of Students to Groups

Students at the two secondary schools were advised there was a study programme available to them and were randomly assigned to one of two study-skills groups, the MS or the TS. Participants in this study knew there were two different programmes while both programmes focussed on study strategies for learning curriculum content and gaining NCEA credits. Students did not choose which study support programme they attended but were allocated in matched pairs with the aim of creating two groups that were as similar as possible. The random allocation of students to groups may be considered equitable to

¹ Decile may be considered an indicator of socio economic status, where one is low and 10 is high; see for example Hattie (2003, p. 7).

all students because both the TS and MS study programmes should have had positive learning effects: no matter which programme the student was allocated to, all students should have received a useful learning opportunity. A variety of pre-tests were administered to all students in both groups before the intervention programme and post-tests after the final session.

Description of the Traditional Study Support Programme (TS)

The TS used in this study had three aims that are similar to those reported by Gorinski (2005): (a) to provide a regular place for students to complete study, (b) to offer one-on-one assistance to students, particularly with literacy activities, but also in other subject areas, and (c) to offer a programme of general study skills. The content of the proposed TS is based on Science or Mathematics and specific study skills and aims to improve student achievement by using examples from the Year 11 curriculum in Science or Mathematics to provide contextual support (Hattie, 2009).

Description of the Motivation-enhanced Study Support Programme

The MS programme incorporated all the components of the TS programme, plus additional, motivation-focused components. These motivation elements come from two main sources: concepts from Martin (2008) and material relevant to the NCEA from Meyer et al. (2006, 2007, 2009) around the motivation orientations *Doing My Best*, *Doing Just Enough*, *Teacher Affiliation*, and *Peer Affiliation*. It is important to note that the incorporation of motivation elements in the MS programme did not diminish the content of the MS programme in comparison to the TS programme. Students in both programmes received practical assistance with their schoolwork especially related to study skills and organisation. The main advantage of operating TS and MS in school time rather than after school was a higher participation rate from students who did not have to make a choice between coming to the programme and attending competing events such as sport or cultural practices or child-care of siblings. This had particular significance for this study which focused on the under-achievement of some Pasifika students. Meyer et al. (2009) noted that “Nearly half of students reported they spent time weekly caring for siblings and other children in the family,” that “Pacific students reported more childcare than others,” and that “Asian and Māori student participation in childcare decreased from Year 10 to Year 11, whereas Pacific students increased childcare across these two years” (p. 99). All MS and TS study sessions were conducted by the researcher who has relevant expertise in secondary education.

Quantitative Data Collection

This mixed-method study was conducted in two phases. In phase one, quantitative data were gathered from a two-group pre-post-test randomised experiment that aimed to determine whether achievement-related outcomes of the two groups were different after each treatment programme. Achievement-related outcomes of students from each group were determined using the six indicators as outlined later. As shown, the design of this experiment can be represented by two lines -- one for each group -- with an R at the beginning of each line which indicates that the groups were randomly assigned to one of two groups MS and TS. Groups A and B were from school A, and Groups C and D were also from school A, and Groups E and F from School B. One group from each of the pairs, A & B, C & D and E & F received the TS and the other group in the pair, which was the relative comparison group, received the MS. Any pre-tests were not for the purpose of group assignment as students were randomly assigned to the two conditions on the assumption that the two groups were probabilistically equivalent. Administration of any pre-tests allowed for evaluation of group equivalence as well as measuring changes in achievement, motivation, and engagement over time.

The research design

Group A: R ----O-----MS -----O
Group B: R ----O-----TS -----O
Group C: R ----O-----MS -----O
Group D: R ----O-----TS -----O
Group E: R ----O-----MS -----O
Group F: R ----O-----TS -----O

R = Random assignment of matched-pairs by day chosen for Study Support

O = Relevant achievement, motivation and engagement measure administered as pre-tests and post-tests (pre-testing was before first study session, all students; post-testing after last study session, all students)

MS = The experimental condition Motivation-enhanced Study programme

TS = The control condition Traditional Study programme

Instruments used in quantitative data collection

Two main instruments were used in this study. The first was Martin's (2001, 2005, 2008) *Student Motivation and Engagement Scale* and the second was Meyer et al.'s (2006, 2007, 2009) *Screening Tool*.

Measures of Achievement-related Outcomes

Measures of achievement-related outcomes used in this study covered four aspects: academic achievement, motivation, engagement, and attendance. Academic achievement at Level 1 NCEA included total credits, credits gained at Achieved grade, credits gained at Merit grade, credits gained Excellence grade, and credits Not Achieved. Measures of motivation, which included a variety of intra-personal and inter-personal motivation orientations, as well as engagement and attendance measures, were pre-tested (at the beginning of the intervention) and post-tested (at the end of the intervention).

Qualitative Data Collection

In phase two, following initial analysis of quantitative data, qualitative data were gathered from interviews with a sub-sample of students from the two schools where the programmes were implemented. Students were interviewed in pairs to ensure students' comfort by being with another student while being interviewed. All interviews were conducted in a public area – the school library – as this was found to be a non-threatening environment that encouraged open responses. While it was noisy at times with other students in the background, this setting was considered more comfortable for students, as interviews were in a neutral place with many other students present. It was planned that the interviews generally comprised either two female students or two male students to avoid any possibility of gender dominance. The researcher interviewed pairs of students from each of the MS and TS programmes at each of the schools until such time as the data were saturated. Interviews were approximately 45 minutes long and were tape recorded and then written down without any students being identified by their names. To ensure comparability between schools, this research was conducted in two state, co-educational, similar decile secondary schools in a matching geographic location.

The interviews were semi-structured but many open-ended questions spontaneously arose. This interview structure approximated a methodology referred to as 'talanoa' which is a qualitative oral approach considered to be culturally responsive towards participants such

as the Pasifika students in this study (Fletcher, Parkhill, Fa'afoi, Taleni, & O'Regan, 2009). Features of talanoa include spontaneous discussion, conversations that flowed freely with less formal interview structures, fewer pre-determined questions, which "reduces the gap between the researched and researchers and gives the researched shared ownership over the direction and focus of the discourse" (Fletcher et al., 2009, p. 26).

This study is firmly contextualised in the lives of students at school and it was important to spend time getting immersed in the school communities. The researcher spent some time at each school before any research was done, attending assemblies and other school events and gradually coming to understand the specific context of each school as the school setting of the participants. This approach raises issues about researcher subjectivity/objectivity. As Morgan (2007) noted that "although one often hears arguments about the impossibility of "complete objectivity," it is just as hard to imagine what "complete subjectivity" would be" (p. 71). A feature of this approach was being able to follow the principle of inter-subjectivity. This relates to the relationship between the researcher and the research process and the ability to build on reciprocal relationships.

After initial analysis of quantitative data there were two phases of interviews. Part One interviews occurred immediately after the intervention programmes and were based on the following questions:

- Tell me about the study programme. Why did you participate? What did you like about it? What didn't you like?
- What part of the programme helped you the most? What part didn't help you or didn't make sense to you?
- Give an example of something you learned in the programme that you used later, in one of your assessments in the classroom? What happened?
- What kinds of things can teachers do to help support you in your classes?
- How can your classmates and friends support you in your classes? Outside school?
- Do you find it helpful to bring your own experiences and/or cultural background into your schoolwork? How?
- Are there any barriers that make it hard for you to achieve at school. What are they?
- How do you think the programme could be improved to make it better meet your needs?

Part two interviews were based on the findings emerging from the quantitative and qualitative analyses:

- Tell me about how you studied for the external exams for Level One NCEA
- How do you feel about school?
- Tell me what makes you want to work hard at school
- Tell me about something you are really good at
- Tell me about how someone had influenced you to work hard? (positive and negative influences).

Data Analysis

Because the quantitative data was not the main aim of this paper, the focus of data analysis was on the qualitative data. After analysis of the quantitative data, qualitative data were collected and analysed with the aim of building on the initial quantitative analysis. Mixing of the data occurred when the initial quantitative data informed the qualitative data collected and were subsequently analysed. In this way, the study followed a sequential explanatory design that aimed to explain and interpret the quantitative results

by collecting and analysing follow-up qualitative data. This was particularly useful to examine the unexpected results that arose from the interviews, especially about how some students used their study time and their misunderstandings about some examination procedures (Creswell, 2009).

This study used grounded theory to analyse the data from interviews in the systematic stages of open, axial and selective coding. The open coding process generated initial categories of information and axial coding facilitated selection of the categories by positioning them into the theoretical perspective of this study. The interconnection of the generated categories into a coherent description was achieved in the final process of selective coding (Creswell, 2009).

Some of the data that emerged from the initial analysis of the qualitative data provided themes investigated in this paper. These were the derived grade process, the equipment students believed they could legitimately request during external assessments, how students spent their study leave time, how parents supervised study leave, and how the school support compared with home support. These codes were positioned in a theoretical framework proposed by Meyer et al. (2009) which outlined three motivational influences on achievement: intra-personal, inter-personal, and external motivational effects. Intra-personal motivational influences include an individual's self-perception of ability or their sense of self-efficacy. Students may also be motivated by inter-personal motivational concepts such as peer-affiliation, teacher-affiliation and or family-affiliation, rather than intra-personal factors such as self-belief. External influences on motivation and achievement could include sport, caring for other children in the family and part-time work. This paper focused on responses from students about self-managed study leave which they considered supported or were a barrier to their study efforts while they prepared for external examinations. In this context, study leave may be considered to be an example of an external influence on motivation and achievement.

Findings

The Derived Grade Process

The data showed some students did not appear to understand the derived grade process. The derived grade process is outlined by the New Zealand Qualifications Authority (NZQA) as a method of establishing grades for NCEA where a student has been prevented from sitting exams because of circumstances beyond their control, for example, a death in the family. Schools use derived grades recorded during the year from valid practice assessments to decide what grades to award to students who have been legitimately absent for external assessments. While NZQA advises schools that students need to be made aware that practice assessments may be used to generate a derived grade, and despite schools informing students about the derived grade process, not all students reported understanding how this process worked. For example, two Year 11 Pasifika students were asked how they used their study leave, and one student (S) reported; "We had a family thing going on at the same time – because we are to um because the week after we finished school my uncle passed away and we had to help out with the family and yeah."

When (S) was asked if he sought a derived grade he did not appear to understand the question even when the researcher gave the following explanation: "If someone in your family passes away and you've got exams you can apply to the school to say you were affected by this and they can give you help... it's called a derived grade." (S) replied, "Oh I didn't know you could do that". His friend (T) concurred that he did not know about derived grades either.

Consequently more students were reinterviewed, which revealed more evidence of misunderstanding about derived grades. Two more Pasifika students explained how they spent their study leave. (P) reported:

I went to where my Nan was — a few days later she got sick so I was just staying with my Dad and she had a heart attack and so we had to go to the hospital and I couldn't go to the ... test kind of thing.

When the researcher asked (P) if he had gone for a derived grade, giving a similar explanation as above, (P) seemed surprised and (P) and his friend (T) discussed this apparently new information; (T) “Oh wow I'd thought that if you miss it” (the assessment), (S) “That's it for the year”, (T) “Yeah and you don't get those credits.”

Equipment Students can Legitimately Request During External Assessments

Other students reported that they could not ask for equipment such as a pen during external assessments. This does appear to be an area where there is some confusion. Information presented to students by NZQA (2011) in a document called *Instructions to Candidates*, states: “You have to take your own equipment such as pens, pencils, calculator and an eraser into the exams. **You may not borrow** (NZQA's emphasis) equipment from someone else during the exam.” Some students appeared to have interpreted this information literally. For example, when two Pasifika students (PT) and (PM) were asked about how the examinations had gone, (PT) replied; “In English I think I failed and you know the Science we get three papers each with five credits. I went in the morning and went over into class (to do the exam) and meantime I had no pen so I sat there for the whole 45 minutes without doing anything.” When asked why he didn't ask for a pen? his friend (PM) explained, “You know they tell us at the start that you can't ask for anything when are in the exam room.” (PT) continued with his story, “Yeah so I left after 45 minutes yeah and... and when asked how he felt about that replied; “Disappointed because I could have done some of the Science in the three and I didn't pass (NCEA).”

Remarkable as this may seem, both (PT) and (PM) seemed convinced that they could not ask for a pen during the examination. Again following the emergence of this finding from the initial interviews, further information about asking for equipment was sought. (S) concurred stating, “They even told us that last year that if you don't have a pen or anything you can't ask the teacher.”

How Students Spent Their Study Leave Time

It is common practice in New Zealand's secondary schools to release Year 11 students on study leave for as much as four weeks of term time prior to the external examinations. However, some students reported that they found it difficult to resist spending study leave time ‘at the beach’ with their friends, rather than focussing on study. The rationale behind study leave appears to be situated in the principle of self-management. The New Zealand Curriculum (Ministry of Education, 2007) explains that self-management skills are one of five key competencies considered important in New Zealand schools and that self-managing students is:

Enterprising, resourceful, reliable, and resilient. They establish personal goals, make plans, manage projects, and set high standards. They have strategies for meeting challenges. They know when to lead, when to follow, and when and how to act independently. (p. 38)

However, evidence from interviews showed variable levels of student self-management. Some students, such as (E) reported successful strategies: “I looked over my notes that I wrote in class and I studied every day for one or two hours, sitting down in my room reading stuff ... books” and “I had like breaks every half hour I would do something else like for 10 minutes and then go back and study.” However, other students reported less successful self-management study strategies such as; “No um I did some at night but then it would all go by the morning” and “I would forget so then I would have to do it again but at other times I just couldn't be bothered doing it.” Data from these interviews showed that for some students study leave was like a vacation: “I just chilled-out, had a holiday cause

like it was the first time I had study leave.” When asked to explain what they did, some students said they “Went to the beach” because “friends rang you up and said let’s go to the beach” and “Oh we had study leave for four weeks but if it was a hot day we just made a plan to go to the beach – we went to the beach quite a lot and that’s what we did.” Students quite freely admitted they were not studying when they were at the beach, and rated their study effort as; “Halfway between I didn’t try at all and I tried a little bit.” Others described their study attempts as being less effective: “I did those last-minute ones – just one day before the exam I re-read all my notes and everything sometimes I get it and sometimes I don’t – I reckon I tried a bit but just low.” These same students also reported “I wish I had studied more.”

How Parents Supervised Study Leave

Evidence from interviews also seems to show that some parents were not able to supervise study leave effectively. (T) reported on the subterfuges she used to fool her parents: “They think you are in your room studying but most of the time I’m just sitting in the front lounge watching TV.” When asked what her parents did about that, (T) replied “Well they tell me off and take the remote and then I mostly sit there and look at the papers and that, and they come in and say are you studying – they think I am studying because I have my papers out but when they walk out I stop.”

Other students reported that parents simply did not know what they were actually doing on study leave. (V) explained that her parents left for work before she got up and as long as she was home when her parents got in it was OK. When her parents returned (V) made sure she was home and when her parents asked: “How did your study go?” (V) would reply “Good thanks”. In fact some days (V) admitted to the researcher that instead of studying, she had been out with her friends. It may not be reasonable for schools to expect all parents to monitor measurable daily outcomes from study in the way that teachers would do as part of a supervised study system at school.

Other evidence from interviews indicated that schools are able to provide support that may not be available in some students’ homes. This was particularly in the area of access to computers, printing and tutorials. Year 11 students are encouraged to download papers from previous examinations from the NZQA website, a free service for students at school. However, once at home on study leave this resource may not always be available. (S) reported: “I can’t print on the computer at home because it’s my uncle’s computer.” Schools can also offer tutorial assistance, when students are physically at school but not as easily when students are on study leave. This assistance appears to be valued by students who noted: “Well we get papers from our teacher – like in English our teacher gave us revision papers – questions that we can do essays on.”

Conclusions

This paper has identified a number of ways students perceive study leave as a support or barrier to their study efforts. This study provided evidence that some students did not understand policies and practices that had a direct impact on their performance. The fact that some students believed they could not legitimately request equipment during external assessments may be interpreted as evidence that they have not listened properly as the teacher gave out examination instructions. This is because other students reported that you could ask for pens, extra paper, and tissues. However, it is also possible that some students have received mixed messages from various sources. This could be teachers warning students to bring their own equipment such as calculators. The NZQA instruction, sheet, which states: “You may not borrow equipment from someone else during the exam”, may give some students the view that ‘someone else’ includes the exam supervisor. Common-sense would indicate that a student is not supposed to sit there doing nothing in an examination because they have no pen, but the evidence from this study shows that some students do interpret the instructions literally. Additional issues could include

students being less familiar with the exam supervisor who is not a teacher at the students' school and does not, of necessity perhaps, understand individual students in the way that teachers who know the students do. Perhaps some students are less sure about asking less 'familiar' people, like the examination supervisor, for help.

Evidence about how students spend their study leave time indicated that perhaps schools should allow only some students to go on self-managed study leave, i.e., those who have 80 credits, including literacy and numeracy requirements. Students without this level of achievement may show they do not have the self-management skills required to study successfully on study leave, and perhaps should be required to stay at school under supervised study. Students most likely to be unable to self-manage, as evidenced by their lack of credits, could be better supported by the school. It is also possible that if schools told students they would not be released for self-managed study leave unless they had 80 credits, this could be a motivating factor for students to work hard during the school year. Managing students who are kept at school should be relatively easy for schools as teachers would have been relieved from their teaching duties because most of their students would have been released on study leave. Such teachers would take responsibility for supervising the underachieving students who had not reached the 80 credits needed for Level 1 NCEA in an attempt to maximise a student's chances of success. This could delay students dissociating themselves from school to prepare well to gain credits from external examinations.

What has been called 'Gone to the Beach' in this study could also be interpreted as representing student disengagement from school. Because students have been released from the organisational structures of school, they may have also moved past being ready for exams, become demotivated, and believed that they are on holiday. It is possible that as the 'school' pattern has been disrupted, dissociation with school has also commenced. Yet credits are still available for students if they prepare for, and attend, external examinations. It may be appropriate to delay this observed dissociation with school by keeping students physically in the school environment as long as credits are still available. Schools may need to consider having a better communication system with students who are on study leave because students may fail to communicate with their teachers about events that are impacting on their motivation and achievement such as bereavement. This may be compared to the situation when students sit internally assessed standards and any absence from school is immediately noted, and if it was for bereavement students may get a re-assessment opportunity. However, when students fail to attend an externally assessed standard there may not be any follow-up until the next school year when teachers may discuss results with students. Since this could be January or February of the following year, it is too late to do anything about it. This problem could be overcome by requiring students on study leave to attend school and absences could be monitored and investigated.

Recommendations

This paper makes the following recommendations for students who are already at risk of under-achievement so they are not further disadvantaged by not understanding policies and practices that may have a direct impact on their performance or have lower self-management skills and find it difficult to resist spending the examination study period 'at the beach.'

Schools need to ensure students understand the derived grade process before students go on study leave. Some students appear to need more explicit instructions about what to do if family bereavement or personal illness impairs their study or attendance at external examinations. Schools may need to follow up students who are absent from external examinations the day they do not turn up, to check reasons for absence.

Schools may need to take extra care to ensure that all students present in every external examination understand they can ask for legitimate assistance such as a pen. This may

mean the NZQA liaison teacher, a familiar re-assuring figure to students, needs to take a more active role as part of the problem may be students are less willing to ask for help from less familiar examination supervisors.

Schools may need to reconsider allowing all students to go on study leave. It may be better to inform students at the beginning of their Year 11 study that only students who have achieved 80 credits by early November – study leave time – will be released into self-managed study leave. The under-achieving students (those with fewer than 80 credits) should remain at school, under the close supervision of their regular teachers.

Schools should consider offering study programmes for those students who may have gained 80 credits but who wish to receive more support than they can get at home. This support could include tutorials and access to computers and printing facilities.

References

- Alton-Lee, A. (2003). *Quality teaching for diverse students in schooling: Best evidence synthesis*. Wellington: Ministry of Education.
- Anae, M., Coxon, E., Mara, D., Wendt-Samu, T., & Finau, C. (2001). *Pasifika education research guidelines: Final report*. Wellington: Ministry of Education.
- Atkinson, J. W. (1957). Motivational determinants of risk – taking Behavior. *Psychological Review*, 64(6), 359-372.
- Bandura, A. (2007). Much ado over a faulty conception of perceived self-efficacy grounded in faulty experimentation. *Journal of Social and Clinical Psychology*, 26(6), 641-658.
- Biddulph, F., Biddulph, J., & Biddulph, C. (2003). *The complexity of community and family influences on children's achievement in New Zealand: Best evidence synthesis*. Wellington: Ministry of Education.
- Bishop, R., Berryman, M., Cavanagh, T., & Teddy, L. (2009). Te Kotahitanga: Addressing educational disparities facing Māori students in New Zealand. *Teaching and Teacher Education*, 25(5), 734-742.
- Cook-Sather, A. (2002). Authorizing students' perspectives: Towards trust, dialogue and change in education. *Education Researcher*, 31(4), 3-14.
- Covington, M., & Omelitch, C. (1980). Ability and effort valuation among failure-avoiding and failure-accepting students. *Journal of Educational Psychology*, 77(4), 446-159.
- Creswell, J. W. (2009). *Research design: Qualitative, quantitative and mixed method approaches* (3rd ed.). California: Sage.
- Dickie, J. G. (2008). *An investigation of sites, uses and practices for literacy in the lives of Pasifika students* (Unpublished PhD thesis). Victoria University of Wellington.
- Dweck, C. A. (1992). The study of goals in psychology. *Psychological Science*, 3(3), 165.
- Fletcher, J., Parkhill, F., Fa'afai, A., Taleni, L. T., & O'Regan, B. (2009). Pacifica students: Teachers and parents voice their perceptions of what provides supports and barriers to Pacifica students' achievement in literacy and learning. *Teaching and Teacher Education*, 25, 24-33.
- Gorinski, R. (2005). *Pasifika Islands School-Parent-Community Liaison Project: Case study*. Wellington: New Zealand Ministry of Education.
- Hattie, J. (2003). *New Zealand education snapshot*. Paper presented at the Knowledge Wave Forum, Auckland, New Zealand.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.
- Hattie, J., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research*, 66(2), 99-136.
- Jones, A. (1991). *At school I've got a chance: Culture/privilege: Pasifika Islands and Pākehā girls at school*. Palmerston North: Dunmore Press.
- Martin, A. J. (2001). The Student Motivation Scale: A tool for measuring and enhancing motivation. *Australian Journal of Guidance and Counselling*, 11, 1-20.
- Martin, A. J. (2002). Motivation and academic resilience: Developing a model of student enhancement. *Australian Journal of Education*, 46, 34-49.

- Martin, A. J. (2005). Exploring the effects of a youth enrichment program on academic motivation and engagement. *Social Psychology of Education*, 6, 179-206.
- Martin, A. J. (2006). A motivational psychology for the education of indigenous students. *Australian Journal of Indigenous Education*, 35, 30-43.
- Martin, A. J. (2007). Examining a multidimensional model of student motivation and engagement using a construct validation approach. *British Journal of Educational Psychology*, 77, 413-440.
- Martin, A. J. (2008). Enhancing student motivation and engagement: The effects of a multidimensional intervention. *Contemporary Educational Psychology*, 33, 239-269.
- Martin, A. J. (2009). *The motivation and engagement workbook* (9th ed.). Sydney, Australia: Lifelong Achievement Group (www.lifelongachievement.com)
- Martin, A. J., & Dowson, M. (2009). Interpersonal relationships, motivation, engagement, and achievement: Yields for theory, current issues, and educational practice. *Review of Educational Research*, 79, 327-365.
- Martin, A. J., Marsh, H. W., McInerney, D. M., Green, J., & Dowson, M. (2007). *The role of teacher-student relationships in high school students' academic outcomes: A construct validity approach*. Paper presented at Annual British Education Research Association Conference, London, UK.
- Meyer, L. H., McClure, J., Walkey, F., McKenzie, L., & Weir, K. (2006). *The impact of the NCEA on student motivation: Final report to the Ministry of Education*. Wellington: Ministry of Education.
- Meyer, L.H., Weir, K., McClure, J., Walkey, F., & McKenzie, L. (2007). *Longitudinal research on the relationship between the NCEA and student motivation and achievement: Final report to the Ministry of Education*. Wellington: Ministry of Education.
- Meyer, L. H., Weir, K. F., McClure, J., Walkey, F., & McKenzie, L. (2009). *Motivation and achievement at secondary school: The relationship between NCEA design and student motivation and achievement: A three-year follow-up*. Wellington: Ministry of Education.
- Ministry of Education. (2007). *The New Zealand curriculum*. Wellington: Learning Media.
- Ministry of Education. (2008). *NEMP*, Retrieved from [http://educationcounts.edcentre.govt.nz/statistics/schooling/what is this??student-2009](http://educationcounts.edcentre.govt.nz/statistics/schooling/what%20is%20this??student-2009).
- Morgan, D. L. (2007). Paradigms lost and pragmatism regained: Methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research*, 1(1), 48-76.
- NEMP. (2008). *National education monitoring report*, 49. Retrieved from <http://educationcounts.edcentre.govt.nz/statistics/schooling/>
- New Zealand Qualifications Authority. (2011). *Instructions to candidates*. Retrieved from <http://www.nzqa.govt.nz/assets/ncea/candinfo-02.pdf>
- Ng, D., McClure, J., Walkey, F., & Hunt, M. (1995). New Zealand and Singaporean attributions and achievement perceptions. *Journal of Cross-Cultural Psychology*, 26, 276-297.
- Phillips, G., McNaughton, S., & MacDonald, S. (2001). *Picking up the pace: Effective literacy interventions for accelerated progress over the transition into decile 1 schools*. Wellington: Ministry of Education.
- Richer, K., Godfrey, J., Partington, G., Harslett, M., & Harrison, B. (1998). *Attitudes of Aboriginal students to further education: An overview of a questionnaire survey*. Paper presented at Australian Association for Research in Education Annual Conference, Adelaide, Australia.
- Robinson, V., & Timperley, H. (2004). *Strengthening education in Mangere and Otara; Final evaluation report*. Wellington, New Zealand: Ministry of Education.
- Rubie-Davies, C., Hattie, J., & Hamilton, J. (2003). *Great expectations: Implications for New Zealand students*. Paper presented at NZARE Conference. Auckland, New Zealand.
- Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review*, 92, 548-573.

- Wentzel, K. R. (1999). Social-motivational processes and interpersonal relationships: Implications for understanding motivation at school. *Journal of Educational Psychology, 91*, 76–97.
- Wigfield, A., Galper, A., Denton, K., & Seefeldt, C. (1999). Teachers' beliefs about former Head Start and non-Head Start first-grade children's motivation, performance, and future educational prospects. *Journal of Educational Psychology, 91*(1), 98-104.

The ORBIT Online Analysis, Profiling and Reporting Tool

Peter Tait, Michael Davison, *Auckland University*

John Hattie, *Melbourne University*

Wendy Kofoed, *Auckland University*

Email: peter@orbit.net.nz

Abstract

Schools can be challenged when considering how to use assessment data in order to raise student achievement. In particular, the dual imperatives of Assessment for Learning as well as annual reporting requirements often act in opposition in schools. This article discusses the resolution to this tension, and explores how schools could make better use of assessment data to contribute to learning processes. It describes the purpose and development of ORBIT, a web-based interactive tool, which was developed for use in schools to support a next-steps approach that focused on the individual student's progress. This multi-faceted tool also fulfils parents' demands on schools for more clarity about how students are achieving. ORBIT encourages progressive (levels-based) rather than comparative (grades-based) use of assessment data in an efficient manner. The development of ORBIT was informed by discussion with school practitioners, academics, business analysts, and designers. Specifically, the authors discussed the qualities required of an analysis and reporting system, and participated in reviewing the tools development and user testing. Feedback from the pilot schools is encouraging. Teachers found that the ability to easily display multiple assessments matched to a common scale with the achievement range shown for each outcome supported moderation processes. It would appear that reporting graphically using levels of achievement over years supports a greater understanding of student progress, goal-setting and self-efficacy. There is likely research potential in regard to student outcomes.

Keywords: *Assessment Data, Progress, Self-efficacy.*

Introduction

Speaking at the NZEI Annual Conference in 2011, Anne Tolley, then-Minister of Education, recognised the need for more support for teachers in making assessment judgments when reporting to parents and boards (Tolley, 2011). In particular, she argued that a nationally consistent approach to teachers' overall judgment (OTJ) of students' progress and achievement would give teachers greater assurance that their judgments are consistent and comparable with other teachers across the country. Although support for teachers in making OTJs (also referred to as 'on balance judgments') is not a newly identified need, the imperative to address this matter has expanded with the introduction of legislation that requires schools to make OTJs in comparison to the National Standards.

As Tolley indicated, schools are required to report progress and achievement to parents in regard to literacy and numeracy, and to provide school level data in school annual plans based on OTJs. The Ministry of Education has provided professional development for teachers to implement the national standards and to develop teachers' understanding of OTJs. However, there are limited tools available to schools that support this need that utilise data held in schools' management systems. ORBIT is a recently developed online tool, currently being piloted in schools, that has much potential in this area.

This article will discuss the tensions between assessment for learning and meeting the demands of national reporting requirements. It will describe the purpose of ORBIT, its development, pilot implementation, and potential expansion.

Overview of ORBIT

The rationale for the development of ORBIT was to provide schools with a web-based tool utilising data typically stored in schools' computer management systems, and to align these data with an Assessment for Learning approach focused on student progress. ORBIT supports teachers in daily decision-making about how to determine 'next learning steps' for their students and promotes self-efficacy. Assessment data already inputted in a school's management system is used for multiple purposes to ensure that teacher workload is not increased. ORBIT currently provides interactive graphic displays of a student's progress in reading, writing, and mathematics over their school years. It includes the facility to add levelled teacher judgments and future goals. The displays are able to be accessed by students, teachers, parents, and school management. These displays meet the need for student-parent-teacher conferencing and as a reporting tool.

The concept for this profiling analysis tool to track student progress was also a response to the identified need for multiple use of data, notably assessment for classroom use and schools' reporting requirements. The major value of using assessment data from multiple perspectives in a progressive way is that it allows assessment for learning benchmarks from which a student's individual learning needs can be targeted, and allows the achievement level attained by students in a manner to provide feedback to improve (Hattie & Timperley, 2007).

Statistical and conceptual development of ORBIT began in 2003 with programming initiated in 2010. ORBIT, named for its cyclical nature, draws on achievement data held in school databases, converts to personalised format, undertakes analysis, and generates individual and group reporting (Figure 1).

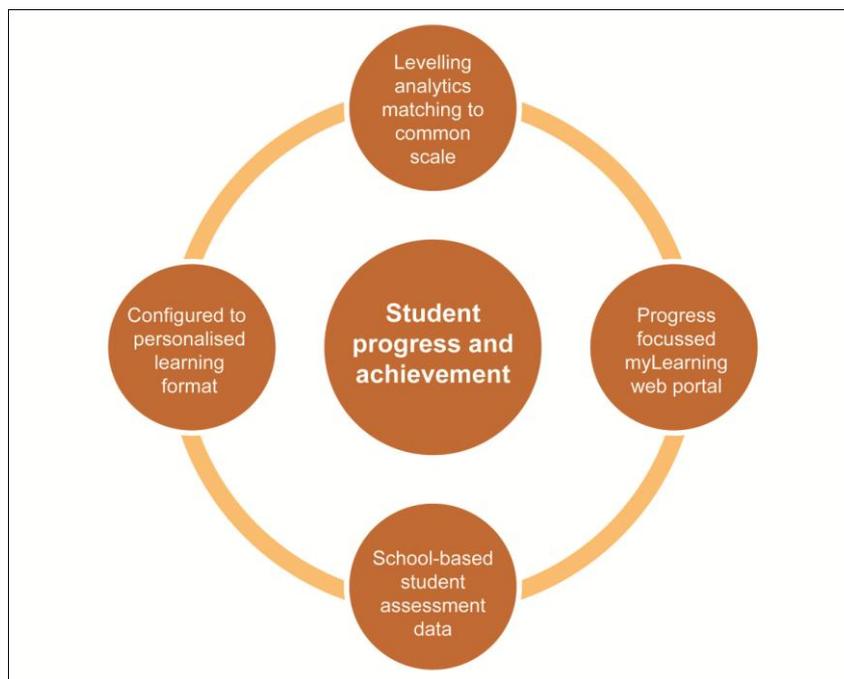


Figure 1: ORBIT Cycle

The use of business analytics enabled the building of a background analysis model with an electronic data management process that transformed typical class-based data into a highly personalised interactive student learning display. For example, the data might include reading ages, Progress and Achievement Test (PAT) scores, STAR raw scores, and asTTle levels. The online tool converts these different results into a common scale. Because ORBIT utilises all data, it indicates the trend which avoids single result variability

or undue emphasis when considering an individual's achievement level. Moreover, ORBIT is able to merge dependable data about student progress to allow interpretations about progress and also to assist teachers in the daily decision-making about how to determine 'next learning steps' for students. A more comprehensive overview of the ORBIT system is available at www.orbit.net.nz.

Background Information

By focusing on individual level progress, ORBIT is able to combine the imperatives of National Standards and the community demand for openness of student results. Research literature shows that 'next steps' use of assessment data supports learning and helps raise student achievement. The key arguments made here are for a progress-focussed use of assessment data, that single assessment events should target individual learning needs, and then contribute to but not dominate the big picture of a student's achievement, supporting new OTJ requirements.

An assessment approach that focuses on a student's individual progress can have a positive impact on student achievement. Crooks (1993) explained that "careful monitoring and reporting of the progress (change across time) of individual students ... is a great challenge to teachers, but offers the prospect of supporting the motivation of all students" (p. 3). Students' views of their progress, in turn, can build self-efficacy. When students succeed on learning tasks, they become motivated to aim higher, and attempt unfamiliar tasks (Pajares & Schunk, 2002).

The myLearning display of an individual's results provides feedback to report and track student learning on a progressive scale. This was based on the well-known Plunket graph – familiar to New Zealand parents. It was decided that a user-friendly graphic display would meet parent and community demands for more openness of student results, and realistic information on how individuals and schools are doing (see Absolum, Geary, McMahan, & Tait, 2003).

The Plunket graph depicts an individual's physical growth over time, and also provides comparative normative information. The emphasis is on the individual's relative growth, and interventions are designed to encourage an optimal pathway (as indexed by the normative information but translated to the trajectory of the individual). A similar Plunket-style graph in education would encourage constructive dialogue between students, parents, and teachers, aimed at developing realistic goals based on the individual's prior achievement. Further, it provides a basis for positive feedback for the student, which is important in building self-efficacy, and improving motivation. Ease of understanding for teachers and students is enhanced by the use of curriculum levels that are well established in New Zealand, and if displayed graphically clearer for parents.

There is an inherent tension in schools – between management's need for data for annual reporting and the pedagogical demand for assessment for classroom teaching purposes. Black and William (1998) noted that "[t]here are clearly difficult problems for teachers in reconciling their formative with their summative roles" (p. 18). There are legitimate purposes for using assessment data, and teachers and school leaders need to see valuable reporting from multiple perspectives. Harlen (2005) observed: "Many teachers have a narrow view of assessment and do not know how to respond to freedom to use evidence from students' actions, projects and processes" (p. 249).

Assessment can provide a way to develop constructive feedback on achievements, to share learning goals and agree on next steps. International assessment literature and the study of teachers' professional practice show how to use assessment data to inform learning and to raise student achievement. For example, in 1999 the Assessment Reform Group produced guidelines for "assessment for learning in practice" (p. 7). These guidelines explain how assessment can improve a pupil's learning when it incorporates

both pupil and teacher reviewing the pupil's work. Research shows that the only way students are successful in reaching their learning target is when they have a clear understanding of their goal, and what is required of them to achieve it (Black, Harrison, Lee, Marshall, & Wiliam, 2004, p. 14).

The value of using multiple assessment sources combined onto a common scale is that it allows assessment targeted to the individual, provides benchmarks from which the student's individual learning needs can be targeted, and allows the reporting achievement level of students in a manner to provide feedback to improve. As Hattie and Timperley (2007) argued, "It is the feedback information and interpretations from assessments" that are imperative, and not the grades or numbers (p. 104).

By modifying the assessment strategy to reduce over-reliance on one-off tests, teachers can focus on individual student achievement and progress, which can promote an increase in teacher job satisfaction. Discussing assessment information with the student and focusing on progress encourages students to think for themselves, and to build their own understanding, and helps to increase their motivation. The students' major concern then is how to improve their learning. To address this concern, the place to start is to look closely at their current level of learning and their journey to that point.

By focusing on individual-level progress, ORBIT is able to combine Assessment for Learning objectives and the request for openness of results – both of which are valuable components in raising student achievement.

Implications for National Standards

Following the announcement of National Standards policy, an increased requirement of schools has been to show evidence of an OTJ at least once a year. This has the potential to reduce over-reliance on single assessment results because increasingly schools are relying on the OTJ to compile school targets. Furthermore, the significance of individual outcomes is lessened, which creates the potential to focus on the overall learning.

The detail on gathering, interpreting, and using assessment information is outlined on the Ministry of Education website, Te Kete Ipurangi (TKI). The process of making an OTJ is described by the Ministry as a 'triangulation' of evidence from multiple sources which should be used to determine an OTJ, as shown in Figure 2 below.



Figure 2: Overall Teacher Judgement 'Triangulation' of Evidence

Further guidance around the way assessment information should be used is outlined. It is expected that students' performances will vary from day to day depending on the nature of the assessment task, the conditions in which the assessment is undertaken, the purpose of the assessment, students' preparation, and the students' engagement and motivation.

ORBIT has been carefully designed to meet multiple purposes and needs of various audiences but especially to provide support for teachers. As William and Black (1998) note, "Standards are raised only by changes which are put into direct effect by teachers and pupils in classrooms" (p. 19). Hence the complete design approach behind ORBIT has been to reinforce and support sound teaching and learning practice.

Common Scale

ORBIT uses existing data drawn from an electronic markbook and repackaged into a personalised learning format where individual assessment outcomes are dated. Analysis then matches a wide variety of assessment outcomes onto a common graphically presented scale. That in turn facilitates a view of learning more consistent with a next-steps approach, and allows workload reduction by offering a moderation tool consistent across schools but informed by local practice.

ORBIT has a robust foundational platform to manage the data flow from multiple schools sourced from two distinct electronic record-keeping databases (eTAP and MUSAC). Through data import protocols, the results are 'standardised' into a common format, transforming data from an electronic class markbook including such logic as imputing dates where required, and merging into a personalised format to meet progress-focused demands. The resultant DataMart is securely managed offline with display information prepared through export protocols to the online application.

ORBIT's approach offers teachers an interface to make more considered OTJs by streamlining the statistical and data management currently done by schools. School trials pre-ORBIT development involved creating and testing a new statistical model that took outcomes from the wide range of existing scales such as Reading Ages, PAT scaled scores, NumPA stages, STAR raw scores, and asTTle levels, and matched them to a common scale. Feedback from many practitioners was that the analysis accurately aligned with their expectations.

Teachers required a solution that was not only accurate but also time-efficient since their own efforts to compare data were inconsistent and enormously time consuming. Hence, the internal programming was checked and enhanced in partnership with statistician Tony Aldridge and analysts (Datamine). Funding was received from the Ministry of Science and Innovation, enabling an accelerated and improved development.

ORBIT assists parent-student-teacher understanding of multiple assessments by displaying them on a single graph. This part of development was undertaken by designers (gardyneHOLT) with an eye to creating a user-interface that would meet teachers' needs. Many teachers are not familiar with web-based applications and need training to use any tool, yet they have limited time to adopt a new system. The in-school testing and modification of ORBIT has been a significant factor in the piloting of interface, now with significant product testing completed and further developments planned.

As displayed in Figure 3, the online tool integrates these different results on a single display with a common scale – represented by grid lines on the myLearning graphic, which eases the difficulty of within-school moderation and provides between-school consistency.



Figure 3: myLearning Assessment Display

Test information is drawn from school databases, including assessments. As these assessments all use different scales, teachers expressed some difficulties in comparing information across these scales in order to make an overall level judgement for each student.

Figure 3 shows a shaded diagonal band which serves the same function as the expected-growth indicator on the Plunket graph. The band matches expectations set by the Ministry of Education regarding typical achievement for students based on their years of schooling. In reference to the left-hand scale the Curriculum Levels 1 to 5 are each a two-year expectation which aligns with the years of schooling indicated on the right.

An extensive discussion and a variety of assessment display concepts were discussed. Boxplot and confidence interval ranges were rejected in favour of the above 'fuzzy' display, as this allows incorporation of factors such as ceiling effects into the myLearning interface without unduly cluttering the image.

Each assessment outcome is represented on the myLearning graphic as vertical bars (fading out at the ends). The intensity of shading indicates significance of a single result. The likely achievement range for each result varies, and this is shown by the vertical stretch of each 'fuzzy' bar. For instance the asTTle assessment dated 5 October 2011 is displayed on the far right of Figure 3, and it can be seen that this score displays much less variability compared to the Probe Reading on either side which has a wider range of achievement indicated.

User Interface

Figure 4 highlights the key features of the main user-interface of ORBIT, the individual myLearning view.

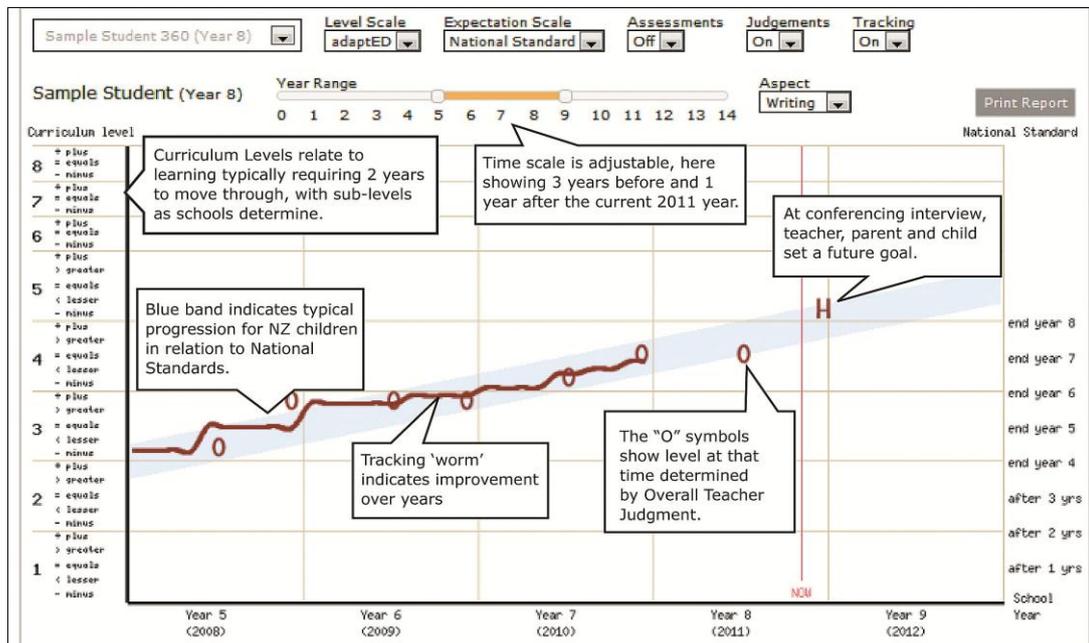


Figure 4: myLearning Interface

ORBIT introduced the concept of students setting goals based on a given level. Kofoed (2009) indicated that parents find goal-setting and 'next steps' useful information to include in reporting (and it is noted that schools are required by the National Standards legislation to provide "What next?" information in reports to parents and students). ORBIT enables goal-setting based on both previous and current data, and these can be used by teachers to set goals that are challenging and achievable. The ability for students to view the history of their progress may provide a positive impact on their learning, specifically because this assessment information provides an indication of their future progress.

To date, ORBIT is the only online progressive tool available that addresses schools' concerns regarding the implementation of National Standards, and does so to reduce the workload around effective moderation.

Figure 5 shows results from different assessments converted to a common scale. The 'fuzzy' bars here are reading assessments. Teachers are requested to set 'best fit' Overall Judgments over years, which are judgements displayed as ovals. In a three-way conference, for example, this student has set a future goal (the goalpost, or H on far right of the graph) which indicates intention to attain the Year 7 Standard by the end of 2011, a suitable goal for this student given past achievement.

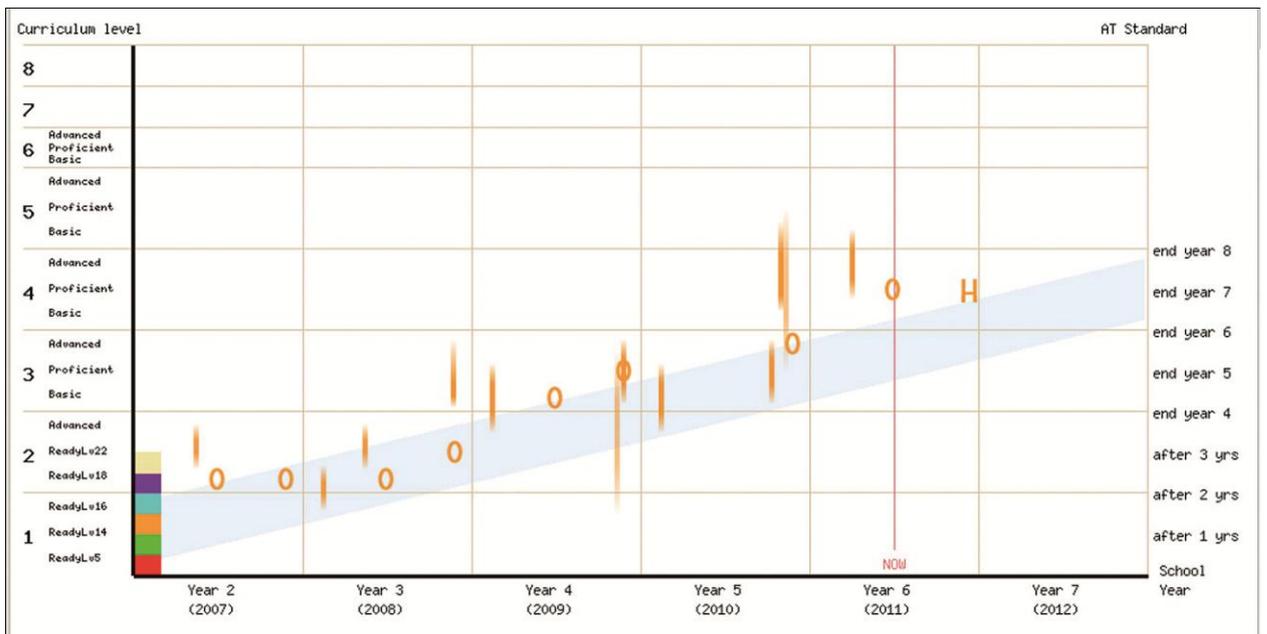


Figure 5: myLearning Student Progress in Reading

The National Standards' required grades (above, at, below, and well below) are to be used in summarising school-wide achievement. These are generated automatically from every overall level judgment made using the myLearning interface, and calibrated individually for each student according to the time of year and the number of months attended at school. The recommendation to schools is that OTJ grades are not shown on individual reports, though displays such as Figure 4 may be suitable for board-level reporting. School leaders also are utilising the Review tool (Figure 6) to support their work in considering improvement targets.

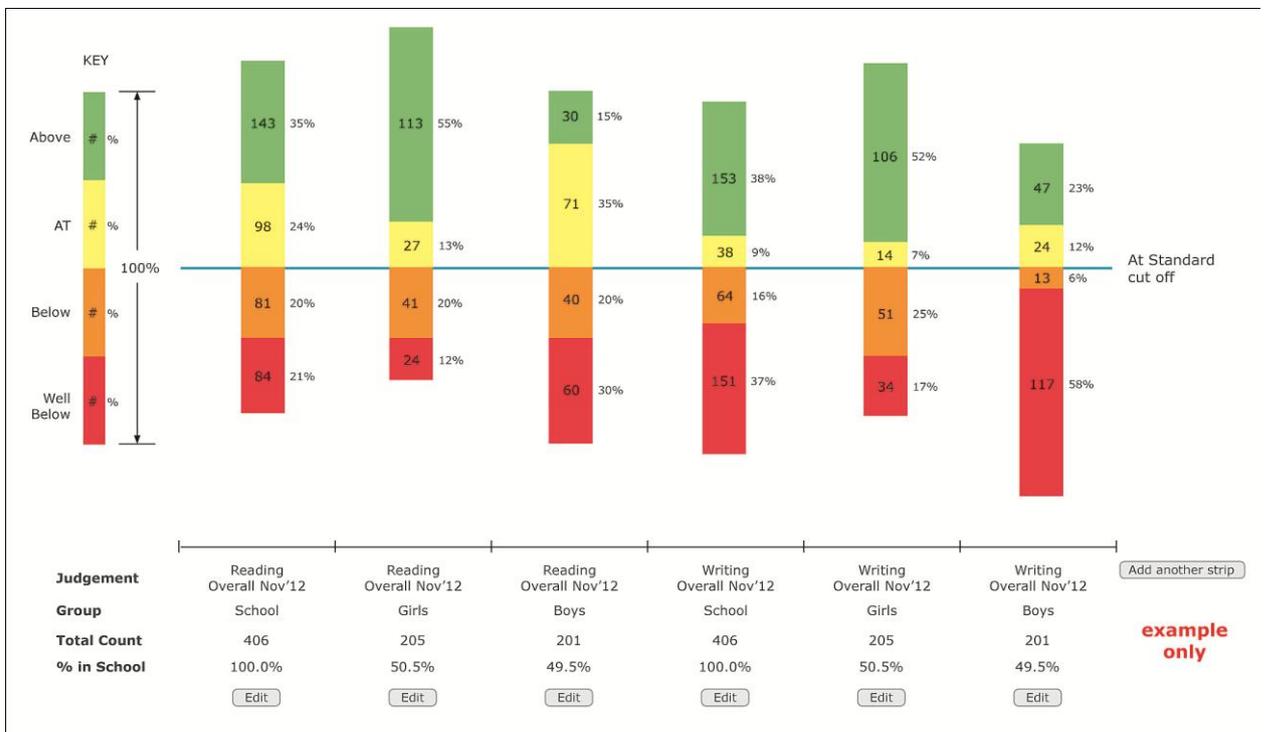


Figure 6: School Review of Reading and Writing OTJ Grades by Gender

Discussion

ORBIT provides teachers with graphical displays of data that demonstrate tracking of results between years. In-depth conversations among teachers revealed the way in which achievement judgments of students are made between a student's teacher of the previous year and their current teacher. Because these data are likely to have on-going use, this has had a positive impact on making judgments justifiable and robust.

ORBIT gathers and integrates assessment data and makes the data easy for teachers to review evidence based on achievement over at least the past 18 months (or longer if desired). Through the myLearning graphical display, ORBIT enables teachers to make Curriculum Level-based Judgements, which can be converted to National Standards Grades. This development has helped schools make better use of their assessment data. ORBIT encourages regular low-stakes assessment by treating assessment as a natural process in monitoring and supporting learning, and, it generates a more reliable data trail for students of their learning, and allows the student to create a journal of their progress.

The 20 pilot schools that participated in the development phase provided many years of student assessment data, as well as valuable feedback on the platform through interviews with a cross-section of staff and parents. In developing ORBIT, consultation has been critical. Development has been well received. Further plans include building progress measures and more targeted profiling tools.

Comments from several schools that are using the ORBIT prototype illustrate this tool's positive impact on teacher practice. The implementation already shows signs of generating positive change in teacher behaviour. It provides timely analysis which supports moderation processes, and it encourages teachers to have in-depth discussions about the progress of individual students from year to year.

Ease of understanding for parents, students and teachers is enhanced by the choice of superimposing the students' results and curriculum-level expectations. Individual progressive-scale tracking of achievement, we argue, is better than reporting achievement in relation to others. This is because it generates constructive dialogue between the student, their parents and their teachers, which helps to set realistic learning goals for the student and provides a basis for positive feedback. These factors are crucial in strengthening student motivation and building self-efficacy.

References

- Absolum, M., Geary, M., McMahon, T., & Tait, P. (2003). Planning for better student outcomes: Analysing student achievement data. Wellington, NZ: New Zealand Ministry of Education.
- Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*. Cambridge, UK: School of Education, University of Cambridge.
- Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*, 96(1), 9-21.
- Black, P., & William, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.
- Crooks, T. (1993). *Principles to guide assessment practice*. Dunedin, New Zealand: Higher Education Development Centre, University of Otago.
- Harlen, W. (2005). Trusting teachers' judgement: Research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers on Education*, 20(3), 245-270.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Kofoed, W. J. (2009). *Written reporting: Strengthening learning partnerships through purposeful reporting* (Unpublished doctoral thesis). University of Auckland.

- Pajares, F., & Schunk, D. H. (2002). Self-beliefs and self-belief in psychology and education: An historical perspective. In J. Aronson, J. (Ed.), *Improving academic achievement* (pp. 3-21). New York: Academic Press.
- Tolley, A. (2011). *NZEI Annual Conference 2011*. Wellington, New Zealand.

The Benefits of Effective Behaviour Screening for Children in Schools and Kindergartens

Gaye Tyler-Merrick and John Church
University of Canterbury

Email: gaye.tyler-merrick@canterbury.ac.nz

Abstract

Early intervention for children with behaviour difficulties can be effective. Intervening early can prevent the development of serious adverse outcomes such as peer rejection, school failure, and later offending. However, there is a disparity between what the research suggests as best practice and what actually occurs in schools and kindergartens. This paper challenges current thinking about the timing of behavioural assessments and describes the elements of an effective screening procedure for children with early onset behaviour difficulties. The implications of a universal screening procedure for school and kindergarten children are also discussed.

Children with elevated rates of challenging behaviour are variously referred to as children with behaviour problems, children with behavioural difficulties, children with behaviour disorders, and children at risk of antisocial development (Church, 2003; Lane, Gresham, MacMillan, & Bocian, 2001; Reid, Patterson, & Snyder, 2002; Walker, Ramsey, & Gresham, 2004). In this report we will refer to the emergence of early onset antisocial behaviour as *antisocial development* since this is the term which is now commonly used by researchers to refer to children who are developing along an antisocial (rather than a prosocial) developmental trajectory.

Regardless of the terminology used, it is clear from the longitudinal studies that it is the children who present with early onset behaviour difficulties who are at greatest risk for the most seriously adverse outcomes during adolescence and adulthood (Broidy et al., 2003; Fergusson, Horwood, & Ridder, 2004; Moffitt, Caspi, Rutter, & Silva, 2001).

Young children who are at risk of antisocial development are characterised by elevated rates of non-compliance to task requests of parents and teachers, hyperactivity, a collection of task avoidance skills, a set of coercive behaviours (e.g., tantrums) that the child uses to get his or her own way (these also change as the child grows older), a set of antisocial reactions to correction and stop requests from other people, bullying, and socially inept attempts to play and converse with peers as the child grows older. Some of these behaviours (e.g., the inability to enter peer groups appropriately and to play appropriately with peers) are viewed by adults as challenging because they reveal delays in social development. Some (e.g., the non-compliance and the tantrums) are viewed as challenging because they are aversive to teachers and parents, and some (e.g., threatening, hitting, and fighting) are viewed as challenging because they are widely defined as antisocial.

Recently, systematic observation of the home and school environments of antisocial and normally developing children has enabled us to identify the key macro-social and micro-social processes that drive antisocial development (Church, 2003; Lahey, Moffitt, & Caspi, 2003; Patterson, 1982; Patterson, Reid, & Dishion, 1992; Reid, 1993; Walker et al., 2004). Large numbers of controlled experiments have confirmed these processes as causes of antisocial development (Advisory Group on Conduct Problems, 2009a; Church 2003).

These new understandings have enabled us to develop a range of demonstrably effective interventions for children at risk of antisocial development, their parents and their teachers. Included under this heading are effective parent training programmes and effective educational interventions which parents and teachers can use in the home setting and in

early childhood and primary school settings (Advisory Group on Conduct Problems, 2009b, 2011; Church, 2003; Eyberg, Nelson, & Boggs, 2008; McMahon, Wells, & Kotler, 2006; Walker et al., 2004). There are now many dozens of well-controlled clinical trials that have demonstrated the power of these educational interventions to halt antisocial development and to accelerate the development of prosocial skills and attitudes. These outcomes stand in marked contrast to the smaller, temporary reductions in antisocial behaviour which result from pharmacological treatments (DuPaul & Eckert, 1997; Purdie, Hattie, & Carroll, 2002).

The Importance of Accurate Early Identification of Antisocial Development

In order to prevent children with challenging behaviour growing up to become antisocial adults, it is necessary for such children to be identified as early as possible, that is, as soon as the first signs of antisocial development begin to appear. Early intervention with at-risk children is critically important for several reasons. First, antisocial habits, once established, are extremely resistant to change (Snyder & Stoolmiller, 2002). Secondly, early intervention with antisocial children, their families and their teachers is much more likely to succeed (and is much less costly) than later intervention (Advisory Group on Conduct Problems, 2009a, 2009b; Church, 2003; Reid, 1993; Walker, Colvin, & Ramsey, 1995).

At the present time our response to the needs of children who have started down the path and are at risk of antisocial development is far too tardy. In the United States of America three times as many 13 year-olds as 7 year-olds are identified with antisocial behaviour (Walker, Ramsey, & Gresham, 2003-2004). Somehow we have become trapped in a situation analogous to that described by Dodge (n.d.):

A man places himself at the bottom of a waterfall, pulling children out of the churning water before the river plunges into a deep, rocky ravine. He hauls them out, nurses their wounds, and calls to others along the riverbank to help him expand his rescue efforts. It occurs to him that if he could go to the top of the waterfall, he might be able to keep children out of the dangerous waters in the first place. But, he thinks, who will save the children who get pulled over the waterfall while I'm climbing to the top of the cliff? Unable to answer that question, he remains – working feverishly – at the bottom of the waterfall saving the children most in danger, wondering how many he can continue to save in this way. (cited in Smith, 2008, p. 2)

One of the factors that hinders the early identification of children who are at risk of developing entrenched antisocial habits is the lack of a standardised behaviour screening procedure suitable for New Zealand use. Without a reliable screening procedure it is not possible to identify the children who should be receiving assistance and it is not possible to identify the specific missing social skills which should be the focus of intervention efforts. Lack of a reliable screening procedure is also hindering our attempts (a) to estimate the number of children whose development is following an antisocial pathway, (b) to measure the proportion of these children who are being detected by current services, and (c) to measure the cost effectiveness of the services that are being provided for these children at each age level.

While several solutions to these difficulties are possible, the one which appears to provide the most accurate information for the least expenditure is a multiple-gating procedure. An example of a multiple-gating procedure with good reliability data is Walker and Severson's *Systematic Screening for Behaviour Disorders (SSBD)* system. The SSBD diagnostic screening involves three sequential steps: (a) teacher referral of any children who meet standardised definitions of externalising and internalising behaviour problems, (b) completion by the teacher of two rating scales for each referred student, and (c) direct observation by a professional other than the teacher of any children whose behaviour exceeds given cut-offs on either of these rating scales (Walker & Severson, 1992a, 1992b; Walker, Severson, & Feil, 1995).

The screening procedure is also consistent with the functional assessment model. Currently in New Zealand, there is a direct observation procedure under development (to be used at the third gate) which will collect information regarding the consequences of children's compliance and non-compliance. This information can then be used for intervention planning at the tertiary level of the Response to Intervention Model.

The New Zealand Situation

Antisocial development can only be identified by examining the child's social response to social demands in social contexts. A standardised assessment of social behaviour requires, therefore, that the assessment be undertaken in a standardised social context. There is only one social context, which all children are required to attend, and that is school. This means that a standardised screening procedure must be one that can be operated *by teachers and with their current level of training*. If the aim is for universal screening, we simply don't have any other options.

The primary aim of the research described in this report was to develop a culturally appropriate and culture-fair multiple-gating procedure for identifying children and youth with severe behaviour difficulties. A culturally appropriate screening procedure was defined as one where both the procedures that are used and the language in which these procedures are described are viewed as acceptable by the main European, Māori and Pacific nation communities in New Zealand. A culture fair screening procedure was defined as one which identifies the same group of children and youth regardless of the cultural community in which they are being raised and educated.

A multiple-gating procedure was selected for development. A multiple-gating procedure is one where measurement reliability is achieved using a three-step procedure in which (a) an adult such as a teacher nominates any children in their care who display certain specified symptoms, (b) the nominating adult completes a standardised rating scale for each child or youth whom they have referred, and (c) an independent specialist completes a clinical assessment or an assessment involving direct observation for those children and youth who meet a certain criterion on the rating scale. Multiple-gating procedures tend to be very reliable because more than one form of assessment is used. Multiple-gating procedures also tend to be very economical because the least expensive procedure (teacher nomination) is used first and the most expensive procedure (clinical assessment) is used last – and only for those children who survive the first two screens.

Technical Challenges Involved in the Design of an Antisocial Development Screen

The research into screening for antisocial development suggests that the design of a diagnostic system suitable for use in New Zealand faces a number of challenges. These include the following.

1. As children with challenging behaviour grow older, they acquire new forms of antisocial behaviour (Advisory Group on Conduct Problems, 2011; Church, 2003; Reid, 1993; Walker et al., 2004). This factor must be taken into account when rating scale items are being written and when cut-offs are being selected during norming of the instruments.
2. Some behaviour (e.g., a low rate of compliance) is much more predictive of antisocial development than others (e.g., being cheeky to teachers). Screening procedures which take this into account are likely to be more accurate than those which do not.
3. Children who engage in high rates of antisocial behaviour are likely to be at greater risk for continuing antisocial development than children who engage in lower rates of antisocial behaviour. It follows that a screening procedure, which ranks children with challenging behaviour according to their level or degree of antisocial development, may be much more useful for intervention planning and resource allocation than procedures which simply classify children as either exhibiting or not exhibiting antisocial development.

4. What a child *cannot* do (their lack of social skills) is as important as what a child can do. Social skills deficits are better predictors of antisocial development than tantrums because they are visible in all contexts whereas tantrums are not. A behaviour assessment instrument that fails to assess positive social skills (or the lack of them) is of little value for intervention planning because it will fail to identify the missing skills that need to be targeted by any intervention (Hosp, Howell, & Hosp, 2003). It is also the case that some disruptive classroom behaviour is simply a function of factors such as boredom, not antisocial development (e.g., Umbreit, Lane, & Dejud, 2004). It is important, therefore, that any behaviour assessment system obtain a measure of both prosocial as well as antisocial behaviour so that the child who obtains high scores for both disruptive behaviour and positive social skills is excluded by the latter from classification as “antisocial”.
5. It is not sufficient for a screening instrument just to be a good predictor. It needs to be a good predictor which has adequate levels of both sensitivity and specificity (Elliot, Busse, & Gresham, 1993). Sensitivity refers to the screening procedure’s ability to correctly identify children in need of further assessment, that is, the ability to avoid false negatives. Specificity refers to the screening procedure’s ability to accurately exclude children who do not need further assessment or specialised intervention, that is, the ability to avoid false positives. During the early stages of a multiple-gating procedure sensitivity is more important than specificity – to avoid under-identification. During later stages, specificity is more important – to avoid stigmatising normally developing children (Walker et al., 1995).

Method

Development of the Antisocial Development Screen (ADS)

The authors worked under contract to the Ministry of Education to develop a diagnostic screening procedure for antisocial development which could be operated by teachers. In 2005/6 we completed the development and piloting of the Antisocial Development Screen (ADS). The ADS is a multiple-gating procedure involving three steps: (a) a teacher referral phase, (b) a behaviour rating scale phase, and (c) an individual assessment (functional assessment) phase (Church, Tyler-Merrick, & Hayward, 2006).

Gate 1: The Teacher Nomination Procedure

To start the referral process, a sample of teachers was given a carefully worded definition of antisocial behaviour and was asked to nominate (refer) any children in their kindergarten or classroom who engaged in behaviour that matched that definition.

Gate 2: The Behaviour Rating Scale

The second gate of the screen is a rating scale which we have called the *Social Development Scale* (SDS). The rating scale has certain features in common with other rating scales such as the Strength and Difficulties Questionnaire but differed from these in several important ways:

- First, the SDS consisted only of items which are known to be highly correlated with antisocial development.
- Second, the wording of each item was more detailed, more contextualised, and written in the idiom of the NZ classroom teacher so that teachers could provide a more accurate response to the question “How often does the child do this?”
- Third, the scale consisted of equal numbers of positive social behaviour items and antisocial behaviour items. Positive social behaviour items were included both as a check on the accuracy of the antisocial behaviour score and to provide information about the missing social skills which would need to be included in any intervention.

- Fourth, all of the items written for the SDS were examined and discussed over several meetings by a cultural reference group including Māori from seven iwi and representatives from the Solomon Islands, Niue, Samoa, Tonga and Tokelau. These discussions resulted in the removal of four items from the SDS item pool.
- Fifth, four versions of the scale were prepared: one for children aged 3 to 4 years, one for children aged 5 to 8 years, one for children aged 9 to 12 years, and one for secondary school students so that the wording of every item was age appropriate.

Each item in the SDS referred to a behaviour which can be observed in the classroom. Based on what the teacher has seen during the past four weeks, he or she is asked to tick one of five options. These included:

<i>Never</i>	You have not seen this child do this during the past month
<i>Occasionally</i>	You have seen the referred child do this a few times during the past month
<i>About half the time</i>	On about 40 to 60% of the occasions when it would be appropriate to behave in this way, this is the way the child has behaved
	<i>Or</i>
	On about 40 to 60% of the occasions when it would be quite inappropriate to behave in this way, this is the way the child has behaved.
<i>Often</i>	This is the option for behaviours which occur with a frequency somewhere between “about half the time” and “very frequently”.
<i>Very frequently</i>	This is the option for all behaviours which the child uses with a frequency which is largely or entirely appropriate, or largely or entirely inappropriate.

The pro-social items were scored from 1 to 5 and the antisocial items were designed to be reverse scored to give a total score (for a perfectly socialised child) of 150. The Gate 1 and Gate 2 activities typically took about 15 minutes per child to complete.

Gate 3: The Functional Assessment

The third gate of the screen took the form of direct observation of the child in at least two different classroom contexts. This was done to attempt to identify the frequency of occurrence of antisocial behaviour (and to begin the task of identifying the most probable reason why the child is choosing antisocial responses over prosocial responses when responding to social demands).

Piloting of the Screening Procedure

Recruitment of the Sample

A sample of 320 children from Decile 1 to 5 Christchurch kindergartens, primary, intermediate and secondary schools were recruited from a single suburb in Christchurch. The selected suburb was ethnically diverse and contained six kindergartens, five primary, two intermediate and one high school. A member of the research team telephoned each kindergarten and school and provided an overview of the project and an invitation to meet. Three of the kindergartens, three of the primary, one intermediate and the secondary school agreed to participate in the project. Two members of the Project Team met with the Head Teacher or Principal of each of these schools in turn to describe the project aims and requirements. If teachers agreed to participate, members of the project team then met the teachers in a staff meeting where they described: (a) the aims of the project, (b) how

to complete the nomination forms on which they identified any children in their class who regularly engaged in antisocial behaviour or who regularly failed to follow instructions, (c) how the matching well-behaved control children were to be identified, and (d) how the rating scales were to be completed for both the nominated antisocial children and the control children in their class. Control boys and girls were selected at random from the classes of the nominated children by asking teachers to select “the next child on the class roll” who was not being referred and who was the same gender as the referred child. In consultation with head teachers and principals, an information sheet and an informed consent request was sent via the kindergarten or school to all the parents/whānau of the children in each participating classroom. The possibility that their child might be selected for direct observation and the right to withdraw from the study was clearly described. Across all the participating kindergartens and schools only two withdrawal requests were received.

Observer Training Procedures

The project team employed and trained five senior postgraduate university students to complete the classroom observations. The research assistants were trained using 2 to 5 minute video segments of children working (and not working) in a variety of classrooms and kindergartens to a 90% inter-observer agreement criterion on each set of categories. They then worked in pairs during their first four observations in the kindergarten/classroom and their performance was reviewed in order to ensure that the 90% inter-observer agreement criterion was maintained in the field.

Procedures in the Schools/Kindergartens

After the nomination forms and rating scales had been completed by the teachers, arrangements were made with each participating classroom teacher for one of the trained research assistants to visit the school and to complete the two 15-minute classroom and peer interaction observations for each of the nominated children and each of the control children at the kindergarten or school. Control children and nominated children were observed together at the same time on a form in two columns. In 27% of cases, a concurrent observation was made by a second observer working independently in order to provide data on inter-observer agreement.

Data collection was scheduled for approximately 40 pairs of children at each of the four age levels: kindergarten, Years 1 to 4, Years 5 to 8, and secondary. Observations at the secondary school were limited to Year 9 and Year 10 classes. During October and November of 2005, the three gates of the SDP were piloted with a sample of 160 3 to 15 year-old antisocial children nominated by their teachers and 160 control children. Altogether nominations, rating scales, and direct observations were completed for 263 children - 131 nominated and 131 control children plus one extra nominated child. In addition, percentage of agreement checks were undertaken for 70 of the observations.

Results

Prior to data analysis, a case review of all available information on each of the 263 children in the sample with completed data (including data from a Critical Events Questionnaire) was used to classify each child as either Antisocial or Not Antisocial. It is this classification which serves as the criterion in each of the analyses that follow.

In the tables below, the following data are presented first for the teacher nominations and secondly for the SDS:

- sensitivity – the number of children who were identified as antisocial by their teachers as a proportion of all the children who met the criterion
- specificity – the number of children who were identified as not antisocial by their teachers as a proportion of all the children who did not meet the criterion

- positive predictive power – the number of accurate nominations as a proportion of the true positive plus false positive nominations
- negative predictive power – the number of children who were selected by their teachers as control children as a proportion of true negatives plus false negatives
- The hit rate, that is, the proportion of teacher classifications that matched the case review classifications.

Table 1 shows the number of agreements between the nominations of the pilot study teachers and the case review classification as Antisocial vs. Not Antisocial. As can be seen from Table 1, the pilot study teachers were quite accurate in their identification of children with persistent behaviour difficulties and missed only 2 of the 116 children in the sample who had been identified as antisocial by the case review.

Table 1: The Accuracy of the Nominations of the Pilot Study Teachers

	Criterion: Antisocial	Criterion: Not antisocial	
Identified by the classroom teacher as antisocial	114	18	Positive predictive power 0.86
Selected by the classroom teacher as a control child using a random selection procedure	2	129	Negative predictive power 0.98
	Sensitivity 0.98	Specificity 0.88	Hit rate 0.92

N = 262, Chi-Square = 191.95, $p < .001$

The Year 5-8 teachers had the highest hit rate (97%). The Year 1-4 teachers and secondary school teachers had similar hit rates (92%) and the early childhood teachers were the least accurate with a hit rate of 88%.

The child scoring closest to the 50th percentile on the kindergarten version of the SDS had been given a total score of 114. Kindergarten scores less than 114 were designated “low scores” and the remainder as “high scores”. The child closest to the 50th percentile on the Year 1-4 version of the SDS had been given a total score of 105. Year 1 to 4 scores less than 105 were designated “low scores” and the remaining Year 1-4 scores were classified as “high scores”. On the version of the SDS for Years 5-8, the 50th percentile lay somewhere between 105 and 112. For this group, a cut-point of 111 was selected and Year 5-8 scores less than this were designated “low scores” while scores of 111 and above were designated “high scores”. For the Year 9 and 10 version of the SDS, the 50th percentile lay between 112 and 114 so a cut point of 113 was selected and Year 9 and 10 scores below 113 were classified as “low scores” while those at or above 113 were classified as “high scores”.

Table 2: The Discriminant Validity of the Social Development Scales

	Criterion: Antisocial	Criterion: Not antisocial	
Number of students below the 50 th percentile on the SDS	115	13	Positive predictive power 0.898
Number of students above the 50 th percentile on the SDS	1	134	Negative predictive power 0.992
	Sensitivity 0.991	Specificity 0.912	Hit rate 0.947

N = 262, Chi-Square = 211.60, $p < .001$

As can be seen from Table 2, the teacher-provided scores on the SDS were highly correlated with the criterion Antisocial vs. Not Antisocial. Overall, classification based on scores above and below the 50th percentile on the SDS matched the classification based on the case review in 95% of cases indicating that, by and large, the SDS discriminates between children who are and are not engaging in elevated rates of antisocial behaviour in the classroom.

The SDS scores provided by kindergarten teachers produced the lowest hit rate (90%). The scores of the Year 1-4 teachers were the next most accurate (95% hits), followed by those of the Year 5-8 teachers (96% hits) and the secondary school teachers (97% hits).

Analysis of the results from the direct observations indicated that they added nothing to the predictive power of the Teacher Nominations and the SDS. Examination of the recording forms and discussion with the observers suggests that the low discriminant validity of the observations was probably due to the fact that the observation period was too short. In addition, many of the teachers were managing their classrooms in ways which left little scope for student misbehaviour over the relatively short period of time in which observations were being made. There is also the possibility that some of the nominated children were aware of what the observers were doing.

Discussion

All of the SDS discrimination measures were greater than the corresponding measures obtained during the latest evaluation of the predictive power of the Eyberg Child Behavior Inventory (Eyberg et al., 2008) using a similarly sized sample of referred and non-referred children. This suggests that, with cut-offs set at the 50th percentile, the predictive power of the SDS is more than satisfactory.

The ability of a screening procedure to avoid false negatives (misses), that is, its sensitivity, is the property which is most important when screening for antisocial development. This is because the cost of failing to identify a child at risk of antisocial development deprives that child of access to important services and this may result in very considerable social costs over the lifetime of the child who was missed. The pilot study data suggest that the first gate procedure had sufficient sensitivity to identify almost all the children at risk of antisocial development in this sample. If checks are made at ages 5, 7 and 9 years, the simple referral procedure which has been piloted has the potential to identify a high proportion of the early onset of antisocial development present in the school population.

There is a price to pay for a high level of sensitivity and that is the referral of considerably more children for an individual assessment than there are children with antisocial development in the school population. This should not be the cause of any concern. Many

of the children who are identified with high scores on the SDS will, in fact, be presenting with relatively serious disruptive behaviour problems – problems which will need to be attended to either by the school or by the Resource Teacher: Learning and Behaviour (RTLB) service. At the second gate the cost of a false positive is the cost of a direct observation at the third gate. However, this is also the main cost of the functional assessment which will be required if the child qualifies for an individual behaviour management plan. Furthermore, this cost is miniscule compared to the cost (over the lifetime of the child) of failing to identify and treat a case of early onset antisocial development.

The third gate, the direct observation, serves two functions: (a) the function of checking the accuracy of the teacher's report on the SDS, and (b) the function of providing data that enable the school (and the RTLB) to distinguish between children whose elevated rates of antisocial behaviour are due to early onset antisocial development and children whose elevated rates of antisocial behaviour are due to other causes. However, attempts to develop a direct observation procedure for the third gate failed during the study being reported here and is being further developed in an additional study that is currently being undertaken.

The Role of Screening in a Response to Intervention Model

Recent advances in the identification and teaching of children with learning and behaviour needs is being undertaken with a Response to Intervention model (RTI) (Brown-Chidsey & Steege, 2010). In the RTI model there are three levels or tiers of intervention. The first tier is universal and is designed for *all* children. If children are identified as at risk of behavioural difficulties at this level they then proceed to Tier II. The second tier provides specific behaviour-based interventions aimed at group or class activities such as The Good Behavior Game and other such group management techniques. The screening procedure described in this report is designed to identify children who will benefit from Tier II interventions. If the data indicate the child is not responding to the secondary level intervention, the child then progresses to Tier III where an individualised behaviour programme is designed for them. Tier III is designed for children who continue to engage in persistent disruptive behaviour and require individualised contingency management plans and social skills training. The screening procedure is not designed to identify children who require Tier III intervention; this is indicated by a lack of response to the Tier II intervention.

Should we screen young children for antisocial development?

There are many who oppose a universal screening programme to identify children at risk of antisocial development on the grounds that it may do more harm than good. It has been argued that any move to introduce universal screening must demonstrate: (a) that the screening procedure possesses a high level of predictive accuracy, (b) that effective treatments exist and can be provided in the numbers required, and (c) that the screening procedure produces less harm than no screening procedure. "It is difficult to see how any system that identifies children as being at high risk of criminality can be anything but stigmatising. Whatever euphemism is offered by politicians, the public will quickly see through it" (Munro, 2006, p. A9).

On the other hand, a compelling case can be made for providing effective interventions for children with early onset antisocial development, their families and their teachers (Advisory Group on Conduct Problems, 2009a, 2009b, 2011; Church, 2003; Walker et al., 2004). It is difficult to envisage a social programme that could conceivably generate a greater return on investment. Of course, all such interventions must be culturally responsive. In New Zealand, as Macfarlane has noted, such interventions need to be "respectful and responsive to kaupapa Māori perspectives, while not being dismissive of the contribution that western science can make" (Advisory Group on Conduct Problems, 2011, p. 42).

The worry is that we will meet the first requirement, that is, develop a screening procedure with excellent predictive accuracy, and possibly even meet the third requirement of being able to demonstrate that we are doing more good for children and families than harm.

However, there remains a very serious risk that we will fail to meet the second requirement, that is, to deliver an effective intervention to the families and teachers of every child at risk of antisocial development we identify.

At the present time, serious attempts are being made by the Ministry of Education to meet this requirement. The first stage of the Positive Behaviour for Learning action plan (Ministry of Education, 2011) includes the Incredible Years Parenting courses, the Incredible Years Teacher courses, and a New Zealand adaptation of School-wide Positive Behaviour Support for selected intermediate and secondary schools. It will be interesting to see whether funding is made available for a New Zealand-wide implementation of these evidence-based programmes and if so, how long it will take to achieve New Zealand-wide coverage.

Acknowledgments

The authors wish to acknowledge the Ministry of Education for funding the 2005/6 pilot project and the schools and teachers who kindly agreed to trial the 2005 and the current procedures. Both projects could also not have been completed without the many parents and children who kindly consented to be part of the project. The authors give thanks to colleagues who gave their time to critique our work.

References

- Advisory Group on Conduct Problems. (2009a). *Conduct problems: Best practice report*. Wellington, NZ: Ministry of Social Development. March.
- Advisory Group on Conduct Problems. (2009b). *Conduct problems: Effective programmes for 3-7 year olds*. Wellington, NZ: Ministry of Social Development.
- Advisory Group on Conduct Problems. (2011). *Conduct problems: Effective programmes for 8-12 year olds*. Wellington, NZ: Ministry of Social Development.
- Broidy, L. M., Nagin, D. S., Tremblay, R. E., Bates, J. E., Brame, B., Dodge, K. A., ... Vitaro, F. (2003). Developmental trajectories of childhood disruptive behaviours and adolescent delinquency: A six-site, cross-national study. *Developmental Psychology, 39*, 222-245.
- Brown-Chidsey, R., & Steege, M. W. (2010). *Response to intervention: Principles and strategies for effective practice* (2nd ed.). New York, NY: The Guilford Press.
- Church, R. J. (2003). *The definition, diagnosis and treatment of children and youth with severe behaviour difficulties: A review of research*. (Report prepared for the Ministry of Education). Christchurch: School of Education, University of Canterbury.
- Church, R. J., Tyler-Merrick, G., & Hayward, H. (2006). *Identifying antisocial development in New Zealand children. Phase 1: Development of the screening procedure*. (Final Report to the Ministry of Education). Christchurch: School of Education, University of Canterbury.
- DuPaul, G. J., & Eckert, T. L. (1997). The effects of school-based interventions for attention deficit hyperactivity disorder: A meta-analysis. *School Psychology Review, 26*, 5-27.
- Elliot, S. N., Busse, R. T., & Gresham, F. M. (1993). Behavior rating scales. *School Psychology Review, 22*, 313-321.
- Eyberg, S. M., Nelson, M. M., & Boggs, S. R. (2008). Evidence based treatments for child and adolescent disruptive behavior disorders. *Journal of Clinical Child and Adolescent Psychology, 37*, 213-235.
- Fergusson, D. M., Horwood, L. J., & Ridder, E. M. (2004). Show me the child at seven: The consequences of conduct problems in childhood for psychosocial functioning in adulthood. *Journal of Child Psychology and Psychiatry, 46*, 837-849.
- Hosp, J. L., Howell, K. W., & Hosp, M. K. (2003). Characteristics of behavior rating scales: Implications for practice in assessment and behavioral support. *Journal of Positive Behavior Interventions, 5*, 201-208.

- Lahey, B. J., Moffitt, T. E., & Caspi, A. (Eds.). (2003). *Causes of conduct disorder and juvenile delinquency*. New York, NY: Guilford Press.
- Lane, K. L., Gresham, F. M., MacMillan, D., & Bocian, K. (2001). Early detection of students with antisocial behavior and hyperactivity problems. *Education and Treatment of Children, 24*, 294-308.
- McMahon, R. J., Wells, K. C., & Kotler, J. S. (2006). Conduct problems. In E. J. Mash & R. A. Barkley (Eds.), *Treatment of childhood disorders* (3rd ed.). New York, NY: Guilford Press.
- Ministry of Education. (2011). *Positive behaviour for learning: Action plan update*. Wellington, NZ: Author.
- Moffitt, T. E., Caspi, A., Rutter, M., & Silva, P. A. (2001). *Sex differences in antisocial behaviour: Conduct disorder, delinquency, and violence in the Dunedin Longitudinal Study*. Cambridge, UK: Cambridge University Press.
- Munro, E. (August 11, 2006). Perils of screening. *The Press*, pp. A9.
- Patterson, G. R. (1982). *A social learning approach Vol. 3: Coercive family process*. Eugene, OR: Castalia.
- Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *A social interactional approach Vol. 4: Antisocial boys*. Eugene, OR: Castalia.
- Purdie, N., Hattie, J., & Carroll, A. (2002). A review of the research on interventions for attention deficit hyperactivity disorder: What works best? *Review of Educational Research, 72*, 61-99.
- Reid, J. B. (1993). Prevention of conduct disorder before and after school entry: Relating interventions to developmental findings. *Development and Psychopathology, 5*, 243-262.
- Reid, J. B., Patterson, G. R., & Snyder, J. (2002). *Antisocial behavior in children and adolescents: A developmental analysis and model for intervention*. Washington, DC: American Psychological Association.
- Smith, S. (2008). *Linking academic and behaviour support: SSC Region XIII RESPONSE to Intervention Conference*. Retrieved from http://www.5.esc13.net/rti/docs/2008Institute/5_linking.pdf
- Snyder, J., & Stoolmiller, M. (2002). Reinforcement and coercion mechanisms in the development of antisocial behaviour: The family. In J. B. Reid, G. R. Patterson, & J. Snyder (Eds.), *Antisocial behavior in children and adolescents: A developmental analysis and model for intervention* (pp. 65-100). Washington, DC: American Psychological Association.
- Umbreit, J., Lane, K. L., & Dejud, C. (2004). Improving classroom behavior by modifying task difficulty: The effects of increasing the difficulty of too-easy tasks. *Journal of Positive Behavior Interventions, 6*, 13-20.
- Walker, H. M., Colvin, G., & Ramsey, E. (1995). *Antisocial behavior in school: Strategies and best practices*. Pacific Grove, CA: Brooks/Cole.
- Walker, H. M., Ramsey, E., & Gresham, F. M. (2004). *Antisocial behavior in school: Evidence based practices* (2nd ed.). Belmont, CA: Wadsworth/Thomson Learning.
- Walker, H. M., Ramsey, E., & Gresham, F. (2003-2004). Heading off disruptive behavior: How early intervention can reduce defiant behavior – and win back teaching time. *American Educator*. Retrieved from http://www.aft.org/pubs-reports/american_educator/winter03-04/early_intervention.html
- Walker, H. M., & Severson, H. H. (1992a). *Systematic screening for behavior disorders: Users guide and administration manual*. Longmont, CO: Sopris West.
- Walker, H. M., & Severson, H. H. (1992b). *Systematic screening for behavior disorders: Technical manual*. Longmont, CO: Sopris West.
- Walker, H. M., Severson, H. H., & Feil, E. G. (1995). *The Early Screening Project: A proven child-find process*. Longmont, CO: Sopris West.

Reading Achievement over the Transition to Secondary School: Profiling a Problem

Aaron Wilson, Stuart McNaughton, and Mei Kuin Lai
University of Auckland

Email: aj.wilson@auckland.ac.nz

Abstract

Like other developmental transitions, the transition from primary to secondary school can mark shifts in trajectories of achievement. This study aimed to examine possible causes for a 'drop' in reading comprehension achievement over a transition from primary to secondary school through a process of 'profiling'. Participants were students transitioning from 33 primary schools into 7 secondary schools in one rural and small town region of New Zealand. Different samples of students were examined to explore different hypotheses. These included two longitudinal cohorts ($N = 887$) and students and their teachers in 31 mathematics, science and English classrooms in secondary schools. Longitudinal analyses of achievement data (Supplementary Test of Achievement in Reading), observations of classroom instruction and student surveys were used to test different hypotheses about the drop in achievement. Differential movement into and out of the region did not account for the achievement patterns. Neither did the immediate effects of the developmental transition in Year 9, or student and family aspirations at Year 10. The drop was associated with features of the instruction in classrooms. The data were used to both explore the nature of the transition and describe the process of profiling the problem.

The Learning Schools Model (LSM) is a systematic evidence-based process for intervening with schools (McNaughton, Lai, & Hsiao, in press) which, like second generation reform models of school change (Raphael, Au, & Goldman, 2009), focuses on a robust and coherent process of collective inquiry using evidence of learning, teaching, and achievement patterns. Judged by the degree and sustained nature of changes, these models can be very effective (Levin, 2008; McNaughton et al., in press; Rowan, Correnti, Miller, & Camburn, 2009).

The LSM draws on two major theoretical proposals to explain how schools might be more effective in meeting valued outcomes. One is the need to design effective instruction from contextualised evidence of teaching and learning. The second involves systematically and collaboratively collecting, analysing and discussing evidence which is acted on to change practices. The research and development process in the LSM has features of a 'performance science' (McNaughton, 2011) and design research (Morris & Hiebert, 2011) approach to educational phenomena and is thereby able to contribute to building knowledge.

The first phase in the LSM, the 'profiling' phase, has this functionality. In this phase researchers and educators collectively generate and refine hypotheses about possible constraints on a particular aspect of school effectiveness. This is done using a variety of sources of evidence of learning and instruction and by drawing on shared research and professional knowledge. It involves joint sessions in which potential plausible hypotheses are raised and tested based on both the research knowledge and school leaders' knowledge (Lai, McNaughton, Amituanai-Toloa, Turner, & Hsiao, 2009). Underlying the process of problem solving is a view that optimal organisational learning involves organisational adaptive expertise that is institutionalised into schools' structures, practices, norms and culture, and that relies on continued and managed inter-dependence with other experts (Bransford, Derry, Berliner, & Hammerness, 2005).

In this study, we report on the profiling phase in a cluster of seven secondary schools in a rural and isolated area of New Zealand. Through earlier research with the elementary and middle schools, we identified an educational problem in student achievement in the secondary schools. This problem was a mismatch between the nationally high reading achievement of the students at the end of their primary schooling (Year 8; 13 year olds) following a three-year intervention with the LSM, and the nationally low student achievement in qualifications at all three levels of the National Certificate of Educational Achievement (NCEA) which occurs three to five years into secondary schooling. By the end of the three-year project, students exiting primary school were, on average, half a stanine (one quarter of a standard deviation) higher on a standardised test of reading achievement than they had been previously, and one stanine higher than national averages, with no ethnicity differences but with a gender difference in favour of girls (McNaughton et al., in press). This was not matched in the NCEA results three years later. NCEA achievement was generally lower than the national average across several indicators, including the highest qualification attained by school leavers, the percentage of the total roll gaining NCEA Levels 1, 2 and 3, the percentage of the roll gaining University Entrance, the percentage of the roll gaining 'Literacy' credits at Level 1, and attainment of external achievement standards in English (Lai, 2009). For example, from 2004 to 2007 the national achievement at Level 1 varied between 61% and 66% of those enrolled at Year 11, but the pass rates in the region were consistently around 10% lower.

We address two purposes in this paper. One is what the profiling revealed about the actual problem facing these particular schools, thereby illustrating the application of the profiling phase of the LSM. The second is what the profiling adds to our knowledge of this developmental transition in New Zealand.

Several hypotheses were considered through the collaborative inquiry. A number were discarded early as clearly inadequate. For example, one explanation relating to the reliability of the primary school achievement results was discarded because the data were collected by an external research team and there had been rigorous quality control in the assessments in both reading and writing (McNaughton et al., in press).

There were four potentially supportable hypotheses based on the research literature and preliminary discussions. These were not necessarily mutually exclusive and the evidence to support or reject them could be used to clarify one or more of the hypotheses. In the following report we describe in turn the four hypotheses, the methods adopted to test the hypotheses, and the results of each test. Different combinations of data were needed to answer different hypotheses. The particular forms of evidence and their data sets, the measures and the participants are described below associated with each hypothesis. These four analyses are followed by a general discussion.

The overall profiling process, which applied to each hypothesis, involved face-to-face meetings and video conference meetings. At these meetings, analyses of student achievement data and analyses of lesson transcripts were used to generate and test the four hypotheses. Four video conferences were held between the research team and the principals and other school leaders between November 2007 and December 2008. At each of these meetings the research team shared their data related to the identified problem and noted the hypotheses raised by school leaders. In August 2008, two members of the team visited each of the seven schools and presented student achievement data collected so far at a series of whole-staff meetings as well as meetings with just the school leadership team to raise, test, and note new hypotheses raised by staff. Each hypothesis raised was then systematically checked and tested, and the results fed back to staff. While we present these in sequence below, in reality these hypotheses were entertained simultaneously and the analyses proceeded in parallel until each had been completed.

Hypothesis One: School Choice Effects

The school leaders raised the possible hypothesis that the 'best and brightest' students left the region at the end of primary school to go to the larger and perhaps more achievement-oriented city schools outside the region. Research evidence does suggest that Matthew effects of this sort can affect vulnerable schools (Ceci & Papierno, 2005). These 'rich get richer and poor get poorer effects' could occur in schools where there is 'flight' of some groups who have more access to resources to choose schooling options. This flight can undermine the average achievement levels at the school because of the effect on the mix of student achievement levels. Related to this was a view that the students who enter the region at 14 years of age would be likely to have lower achievement levels. Previous LSM research – although in urban schools - had shown that new students entering schools engaged in schooling improvement had systematically lower achievement levels (Lai et al., 2009).

Despite this widely held hypothesis, the primary and secondary schools had not collaborated to collect systematic evidence of students' school choices, and testing this hypothesis would require longitudinal data over the transition.

Method

Exploring hypothesis one required longitudinal data over the transition. Several data sets and cohorts were used to increase the robustness of the hypothesis testing. The data sets, which included primary school data, comprised New Zealand European and Māori students from all the primary or area schools ($n = 34$) in the rural and small town region. The data set allowed us to track two cohorts of students to their local secondary or area schools over a 12-month time period from the beginning of Year 8 to the beginning of Year 9, either in 2007-2008 (Cohort 1) or 2008-2009 (Cohort 2).

The primary schools used a standardised reading comprehension test to assess students at the beginning (February) and end (November) of each school year – the Supplementary Test of Achievement in Reading (STAR) (Elley, 2001). STAR provides standardised measures of reading comprehension using stanines, a nine point scale based on a standard distribution of national normative data. All secondary schools were initially using STAR, apart from one which was therefore not included in Cohort 1's Year 9 achievement results. This excluded 144 Year 9 students from the Cohort 1 analyses of achievement.¹ However, these students' Year 8 achievement was included in the demographic information. The school was included in all results for Cohort 2 as they began using STAR for Year 9 students in 2009.

Using these data sets we could examine the school choice hypothesis: whether the students who were leaving the region at the end of Year 8 or entering the region at Year 9 were characteristically different to those who continued on to secondary or area schools. The analysis comprised all Year 8 students at the end of 2007 (Cohort 1, $n=428$) and 2008 (Cohort 2, $n = 459$). These students were tracked *forward* to their Year 9 schools and divided into 'Leave' and 'Continue' depending on whether they remained in the area or not. New students entering the schools at the beginning of Year 9 could be checked for whether they were characteristically different to those coming from schools in the region. This analysis comprised all Year 9 students at the beginning of 2008 (Cohort 1, $n = 390$) and 2009 (Cohort 2, $n = 407$). These students were tracked *back* to their Year 8 schools and divided into 'New' and 'Continue' depending on whether they came from outside the region or from schools within the region.

¹ We tracked the achievement of that school using the reading test that the school provided (asTTle Reading), (Hattie et al., 2004) and examined the means at national distributions from both tests. The school followed a similar achievement pattern to other schools in Cohort 1.

Results

The first hypothesis proposed that higher achieving students chose to leave the region and hence the mismatch between primary achievement and secondary achievement patterns reflected differential enrolment.

Error! Reference source not found. Table 1 shows the number of students (by gender and ethnicity) who continued studying in the region's secondary and area schools compared with those who left the region after Year 8. The large majority of students (86% from both Cohorts 1 and 2) remained after Year 8. A slightly larger proportion of males left in both cohorts. The percentages of Māori and New Zealand European students who left reflected the proportions in the overall school population.

Table 1: Number of Continuing and Leaving Students by Gender and Ethnicity in Cohorts 1 and 2

Cohort		Continue	Leaving
Cohort 1			
Gender	Male	184	32
	Female	185	27
Ethnicity	NZ European	297	48
	NZ Māori	60	8
	Other Ethnicities	12	3
Total		369	59
Cohort 2			
Gender	Male	191	37
	Female	203	28
Ethnicity	NZ European	304	55
	NZ Māori	67	7
	Other Ethnicities	23	3
Total		394	65

The achievement of students who continued in the region compared with students who left at the end of Year 8 is shown in Table 2. This analysis excludes any student without a test score. The achievement levels of those leaving were generally higher (about six months in reading achievement). There was no statistically significant difference between the two groups in Cohort 1. In Cohort 2, the difference reached statistical significance ($p = 0.048$). Notably, achievement levels of the continuing students were still about one stanine above national expectations (or about a year in advance in reading achievement).

Table 2: Mean Stanines for Year 8 Students who Continue and Leave in Cohorts 1 and 2 at the End of Year 8

Cohort	Continue			Leave			Difference			
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
Cohort 1	327	5.87	1.93	49	6.39	1.86	1.76	374	0.08	0.27
Cohort 2	372	6.00	1.86	63	6.49	1.64	1.98	433	0.05	0.28

We then examined whether the students in Year 9 coming from schools outside the region were characteristically different to those continuing. This analysis comprised all Year 9 students at the beginning of 2008 and 2009. The numbers of students in this analysis included all the students in the Cohort who were now Year 9 in the secondary and area schools (not including any that left the area), plus any new students in Year 9.

Table 3 shows the number of students who continued studying in the region (i.e., the subset of Cohort 1 who continued) compared to those who entered the region at Year 9 (i.e., new students). On average, very few students entered the region at Year 9 (Cohort 1: 5%, Cohort 2: 3%). Both cohorts followed similar patterns with regards to gender and ethnicity.

Table 3: Number of Continuing and New Students by Gender and Ethnicity in Cohorts 1 and 2 in Year 9

Cohort		Continue	New
Cohort 1 (Continue) + New			
	Gender		
		Male	184
		Female	185
	Ethnicity		
		NZ European	297
		NZ Māori	60
		Other Ethnicities	12
	Total	369	21
Cohort 2 (Continue) + New			
	Gender		
		Male	191
		Female	203
	Ethnicity		
		NZ European	304
		NZ Māori	67
		Other Ethnicities	23
	Total	394	13

The achievement for these students at the beginning of Year 9 was, on average, at national expectations for both groups of students. There were no statistically significant differences between those who were new and those continuing in either cohort (Table 4).

Table 4: Mean Stanines for Year 9 Students (Continue and New) in Cohorts 1 and 2 at the Beginning of Year 9

	Continue			New			Difference			
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>	<i>d</i>
Cohort 1	223	6.00	2.04	7	6.14	2.19	0.19	228	0.85	0.07
Cohort 2	373	5.86	1.99	12	5.58	1.16	0.47	383	0.64	0.17

The overall patterns in these results indicated that there was a small but detectable school choice effect operating relating to students who left the region. However, neither the numbers of students leaving or newly arriving, nor the achievement levels of those continuing, could account for the large difference between the high Year 8 scores and the secondary school scores at Year 11.

Hypothesis Two: Transition Effects

Developmental transitions are points at which children’s learning is particularly vulnerable. There is consistent evidence that the transition to secondary school is associated with a change from more intrinsic to more extrinsic motivation for academic achievement, and attitudinal shifts away from achievement (Braund, 2008; Galton, 2009). There are consequent detrimental effects on achievement patterns (Benner & Graham, 2009; Wang & Pomerantz, 2009).

A possible hypothesis was that the immediate transition had a marked negative effect on these students that may not be idiosyncratic to the region but an exaggerated version of a more generalised phenomenon. To test this hypothesis would also require longitudinal evidence over the transition from primary to secondary school. Evidence to do with students’ beliefs about achievement, motivation and attitudes to school would also be important.

A second version of this hypothesis was that any transition effects might be delayed and that students may have experienced attitudinal and motivational difficulties over time, for example as high stakes assessments for national qualifications loomed in the third year of secondary school. The longitudinal research evidence from the New Zealand Competent Children Competent Learners (CCCL) study (Hogden, Ferral, & Dingle, 2009) did not directly suggest this. The CCCL study showed attitudes were variable across the primary years, more so than achievement, but became very stable after the transition to secondary school. However, McGee, Ward, Gibbons, and Harlow (2004) suggested that the transition functions like a process which over time may exacerbate shifts in attitude and achievement for some students. Knowing the longer-term pattern in this would also help disentangle an immediate transition hypothesis from possible hypotheses about the quality of instruction in Years 9 to 13.

Method

The analysis for hypothesis two used the same data bases as those used for hypothesis one and addressed the question of immediate and delayed transition effects, and sought to identify the achievement patterns at secondary school in Years 9 and 10. For the immediate form of the transition hypothesis we examined the achievement of students who sat three reading comprehension tests using STAR at the beginning and end of Year 8, and the

beginning of Year 9 on entry to secondary school. When only those students were included, this yielded 192 (Years 2007 - 2008) and 336 (Years 2008 - 2009) students.

Further achievement data were used for the delayed transition hypothesis. The secondary schools had begun to use a new tool in reading comprehension to supplement the STAR assessments. The new test, Assessment Tools for Teaching and Learning (asTTle) Reading (version 4) (Hattie et al., 2004) was designed to reflect curriculum levels and was seen as providing more direct curriculum-related evidence for secondary schools.

The asTTle Reading scores of Year 9 and 10 students at the beginning and end of 2008 were analysed in two ways: for all students in five schools who sat asTTle Reading tests at the beginning and end of 2008, and for students in three schools who were assessed at both time points. In this profiling phase we collected the data that schools provided, and as such the numbers in each analysis varied according to what the schools had collected.

Results

The second hypothesis was that there was either an immediate or delayed effect of the transition to secondary schools. The reading achievement of students who sat STAR tests at the beginning and end of Year 8 and the beginning of Year 9 is presented in Table 5 and Figure 1.

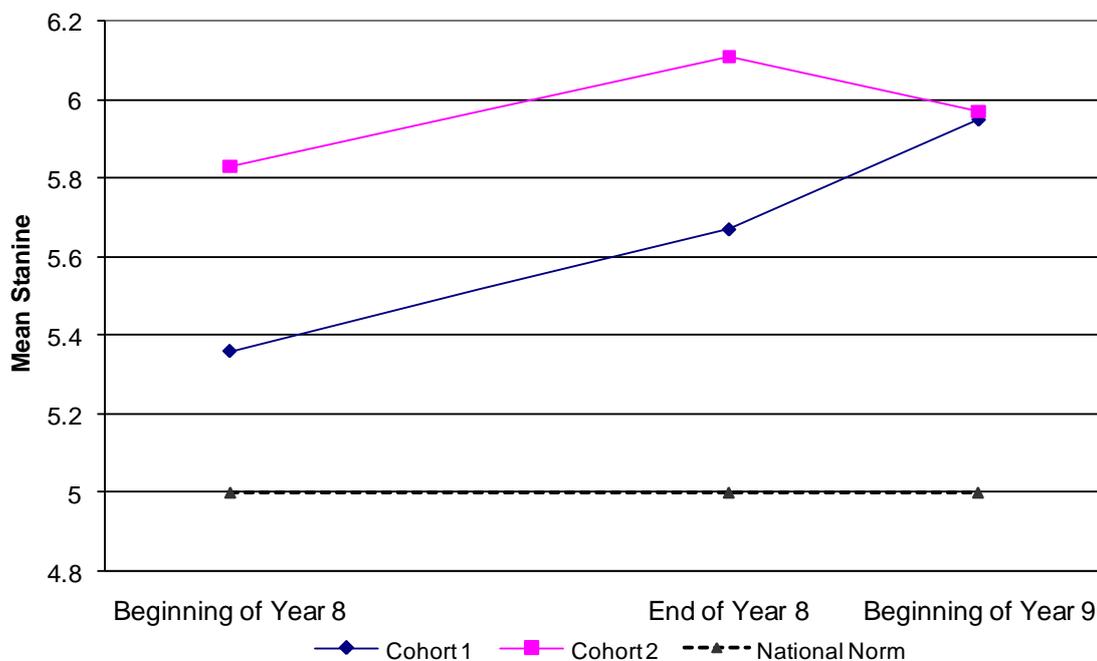


Figure 1: Mean Stanines for Students Who Sat All Three Tests in Cohorts 1 and 2

Table 5: Mean Stanines (Reading Comprehension) for Cohorts 1 and 2 Year 8 to Year 9

	Mean Stanines			T	
	Beginning of Year 8	End of Year 8	Beginning of Year 9	Beginning of Year 8 - End of Year 8	End of Year 8 - Beginning of Year 9
Cohort 1 (n = 192)	5.36	5.67	5.95	4.53***	4.78***
Cohort 2 (n = 336)	5.83	6.11	5.97	4.99***	-2.59**

* $p < .05$. ** $p < .01$. *** $p < .001$.

Achievement in reading comprehension increased significantly from the beginning to the end of Year 8 in primary schools for both cohorts (Table 5). A transition effect or possible summer effect was apparent for Cohort 2; the drop was less than 0.2 of a stanine to around stanine 6. This drop was not apparent for Cohort 1, who made a significant increase over this period. Scores for both cohorts were very similar at the beginning of Year 9 (around stanine 6).

The differences in Cohorts 1 and 2 might be due to variations between schools. Further analyses indicated that for Cohort 2, the drop in achievement was only statistically significant in two of the seven schools. These were large schools at deciles 4 and 6, the latter of which had a larger than the average Māori population (24%).

A more delayed effect of the transition might have occurred, and to test this we examined the achievement data from all students who sat asTTle Reading tests in 2008. Scores are included for 563 students from four schools at the beginning of 2008, and for 389 students from five schools at the end of 2008. It is important to note that students at the two time points presented in Table 6 have not been matched. The Year 9 students at the beginning of 2008 achieved above national norms, but the Year 9 students at the end of 2008 were below national norms. Year 10 students achieved below national norms at the beginning of 2008, then achieved even further below national norms at the end of 2008.

Table 6: Comparison of Beginning and End of 2008 asTTle Reading Scores (asTTle) for Year 9 and Year 10 Students in 2008

	Beginning of 2008				End of 2008			
	<i>N</i>	Cluster Mean	National Mean ^a	Difference	<i>n</i>	Cluster Mean	National Mean	Difference
Year 9	316	560	517	43	172	619	634	-15
Year 10	247	586	634	-47	217	650	728	-77

^a The national norms for Year 9 at the beginning of the year are the Year 8 norms for the end of the year. The national norms for Year 10 at the beginning of the year are the Year 9 norms for the end of the year.

A separate analysis was conducted for only those students who sat the asTTle tests at both the beginning and end of 2008 (matched pre-post students; *n* = 126 Year 9 and *n* = 92 Year 10). The overall pattern for matched students was similar to that identified for all (non-matched) students. This is illustrated in Table 7.

Table 7: Comparison of Matched Pre-post asTTle Reading Scores for Year 9 and Year 10 Students in 2008

	Beginning of 2008				End of 2008			
	<i>n</i>	Cluster Mean	National Mean ^a	Difference	<i>n</i>	Cluster Mean	National Mean	Difference
Year 9	126	581	517	64	126	610	634	-24
Year 10	92	596	634	-38	92	660	726	-68

^a The national norms for Year 9 at the beginning of the year are the Year 8 norms for the end of the year. The national norms for Year 10 at the beginning of the year are the Year 9 norms for the end of the year.

These data showed that the longer students were at secondary school, the further their achievement dropped against national norms, supporting the hypothesis that the causes for this drop relate to students' ongoing experiences of secondary schooling. Further

evidence relating to this hypothesis was collected from the students and used to check both this hypothesis and hypothesis three.

Hypothesis Three: Socialisation Effects

A related hypothesis generated by the schools included the suggestion that education was not highly valued by the community, and that students did not have high aspirations for further education. The degree to which parents value education in general and take an active interest in their own children’s education has been identified as a factor that can impact on students’ achievement and engagement at school (Biddulph, Biddulph, & Biddulph, 2003; Callanan, Kinsella, Graham, Turczuk, & Finch, 2009; Levin, 2008). Local knowledge at the schools was that boys in particular left school early because there was close to full employment in occupations associated with mining, fishing, and timber which did not require further education. In addition, there were, reportedly, historical patterns of growing up which promoted and sustained the expectation that getting a job early was more desirable than completing higher levels of schooling (McGee et al., 2004). The test of this could come from evidence from students themselves on their own aspirations as well as their perceptions of their parents’ aspirations.

Method

In order to test the hypotheses related to socialisation, we administered a survey to Year 10 students ($n = 93$) and interviewed a sample of students ($n = 18$) between the beginning of Year 9 and 10 (O’Brien & Lai, 2011). The survey covered key themes identified by the literature that could influence student achievement: parental expectations and interest in schooling (e.g., Callanan et al., 2009); school and peer group influences such as the sense of belonging at school (e.g., Levin, 2008); students’ relationships with teachers and teachers’ expectations (e.g., Bishop, O’Sullivan, & Berryman, 2010); students’ personal aspirations for the future (e.g., Hong & Ho, 2005) and the pedagogy in use in their classrooms (e.g., Hattie, 2009). Example questions are: “School qualifications are important for the job I want to do” and “I plan to leave school when I am 16.” Each theme was assessed by a series of questions requiring a response on a four-point Likert Scale.

Results

Analysis of student surveys and interviews examined the hypothesis that low community aspirations would sufficiently account for the low NCEA achievement. Mean scores for each theme were generally high ($M = 3.04$ where 1 = negative attitude or perception and 4 = positive attitude or perception) suggesting that students felt positively about each of these themes (Table 8).

Table 84: Mean Responses for Each Theme (where 1 = Negative Attitude or Perception and 4 = Positive Attitude or Perception)

Theme	<i>N</i>	<i>M</i>	<i>SD</i>
Theme 1 – parents	93	3.32	0.46
Theme 2 - teachers	93	2.95	0.54
Theme 3 - peers	93	3.29	0.42
Theme 4 - aspirations	93	3.05	0.64
Theme 5 - engagement	93	2.76	0.51
Theme 6 - assessment	93	2.85	0.45
Mean across all questions	93	3.04	0.37

In other words, students perceived their parents were interested in their schooling and had high expectations of them, that the school and peers were similarly positive about schooling, that the teaching pedagogy they experienced was generally adequate, and that their relationships with teachers were positive. Students also had high personal aspirations and saw school as relevant in supporting them to achieve their aspirations. However, students did highlight one or two areas for improvement. These included a perception when probed in interviews of a lower level of teacher expectations and a lack of academically focused teacher-student relationships, suggesting that rather than a general socialisation effect, improved pedagogical approaches, including a greater degree of challenge in lessons and greater amounts of task-focused feedback, may contribute to the raising of student achievement in the region. This provided more support for hypothesis two and directed attention to possible mechanisms operating in the schools.

Hypothesis Four: Instructional Effects

The fourth general area of hypotheses concerned classroom instruction. This is related to the hypothesis of transition effects outlined earlier because in some research reports, there are descriptions of pedagogy at secondary schools that is 'impoverished' compared with primary schools (Galton et al., 2009). But in the case of these schools there may have been a specific problem related to the quality of instruction. This explanation had been favoured by an Education Review Office (2003) document that stated the overall quality of education in these schools ranged from average to unsatisfactory with considerable variability. There are claims and evidence presented in that report which are contentious and have been challenged. But the report had the effect of raising concerns by the leaders about the evidence for effective teaching.

This hypothesis could take several forms. The first form of this hypothesis was that the instruction might have been of a generally poorer quality. There is considerable evidence suggesting that after taking into account the impact of students' background characteristics, it is what teachers do in their teaching that makes a substantial difference to achievement (Hattie, 2009). This could be tested through analyses of elements of classroom instruction, such as the appropriateness of texts and tasks, and the extent to which instructional time was effectively used. Given that student achievement is enhanced when students are provided with relatively more 'deep' than 'surface' tasks, and that expert teachers focus more on the former (Hattie, 2009), we reasoned that a prevalence of low-order, closed questions would be less than optimal in preparing students for NCEA. Such a pattern might constitute evidence of lower teacher expectations for student achievement.

Similarly, instruction might not be effectively matched to the needs of the students, in particular, if it is pitched at a level that is not sufficiently challenging for students. The researchers developed this hypothesis because initial discussions revealed a widespread perception of student achievement as being lower than the achievement data from the primary schools indicated was possible.

Another dimension to the quality of instruction is teacher-student relationships. There is consistent agreement around the significance of positive relationships with teachers for secondary students (Benner & Graham, 2009; Jia et al., 2009). The significance of teacher relationships, including caring and high expectations, has been identified as particularly important for Māori students and this has support from the results of the intervention programme, Te Kotahitanga (Bishop, Berryman, Powell, & Teddy, 2007). This might have applied to all the students in the region but could be especially significant for the 16% of the students who identified as Māori.

Finally, a very specific form of this hypothesis was that student achievement in national examinations would be enhanced by effective literacy instruction; therefore, less than optimal literacy instruction might contribute to less than optimal student achievement.

There is a large body of research that supports the notion that although different content areas place quite specific literacy demands upon students, the literacy aspects of subjects are not often taught (Draper, 2008; Jetton & Alexander, 2004; McDonald & Thornley, 2005). To test the hypothesis, lessons could be analysed in terms of the frequency and quality of instruction specifically for literacy.

Method

The overall problem of low achievement results in the national examinations was investigated using the data from national databases. The national assessment system at secondary schools (Years 9 - 13), the NCEA, is a modular system that has different units of assessment at three different levels. Some standards are externally assessed and some are internally assessed and externally moderated. Data about individual schools' achievement rates are publicly available online.

To understand how instructional factors may have contributed to the student achievement on NCEA, two sorts of analyses were undertaken. One involved checking the NCEA data. Indirect indicators of the quality of instruction might be found in patterns in the pass rates overall, in internal versus external success rates, and in specific standards. We examined pass rates for one core literacy strand in the NCEA Level 1 exam: achievement Standard 90057, "Read and show understanding of unfamiliar texts." This is externally examined and is an indicator for effective literacy levels, and taps the sorts of skills assessed in STAR and asTTle.

More direct measures came from classroom observations conducted in October, 2008. Two members of the research team observed 31 lessons (15 English, 6 mathematics, 6 science and 4 social science) in the seven schools. All lessons were videotaped and transcribed. The lessons were coded using NVivo software. We analysed the ratio of open to closed questions asked by teachers, and the number of elaborations they requested in response to students' answers. A closed question was defined as requiring a 'right or wrong' single word or phrase answer; a question was coded as open if it required anything more than that. High teacher expectations might be indicated by a high ratio of open to closed questions, and by frequent requests for students to expand and elaborate their answers. A second way of testing this hypothesis was through analyses of literacy instruction. The number of instances per lesson of explicit literacy instruction was checked. An instance of explicit literacy instruction was defined as any informative explicit teacher utterance about reading, writing or vocabulary. The coding was designed to be as low inference as possible and was completed by one coder. A number of the transcripts were checked by a second coder, and any disagreements were settled by consensus. The data that were used in the joint sessions comprised both quantitative and actual transcript data, so there was also collective agreement on the interpretations.

Results

Whether instruction might have been of a generally poor quality was tested through analysis of elements of classroom instruction, such as the appropriateness of texts and tasks, and the extent to which instructional time was effectively used. Each observer was well experienced in classroom observations and making judgements about quality. One of the research team was a curriculum expert in secondary content area literacy. In terms of overall quality and in terms of task and text difficulty, we could propose at least tentative judgements for the group to consider.

We discounted the general form of this hypothesis. Observations in classrooms showed what we considered to be instances of highly effective teaching practice, as well as instances of less effective practice, but overall an acceptable level of quality as judged by curriculum expectations and professional judgment. For example, in almost all cases, the textbooks we observed being used in mathematics, science and social studies classrooms were used at the year level recommended by the publisher. A more specific version of this hypothesis

tested through classroom observation was that instruction might not be effectively matched to the needs of the students; in particular, if it is pitched at a level that is not sufficiently challenging for students. One way we tested this hypothesis was by analysing the ratio of open and closed questions asked by teachers. We found that 75% of all questions asked were closed questions, and that there were relatively few teacher requests for students to elaborate on answers given (56 teacher requests for elaboration in relation to a total of 427 student answers). The hypothesis of lower level challenges was retained.

Another method for checking lower expectations was to examine the standards in which students were enrolled. If it is assumed that enrolling students is influenced by school guidance and expectations, and there is some New Zealand evidence for this (McKinley et al., 2009), then a proxy for general expectations would be the types of standards in which students were enrolled. To test this we identified, by consulting subject experts, a set of externally assessed Level 1 achievement standards that had high literacy demands. Figure 2 shows that the percentages of students from the schools that had been enrolled in these standards were noticeably lower than national enrolment rates; from 2004 there was a downward trend to below a 10% difference from national parentages.

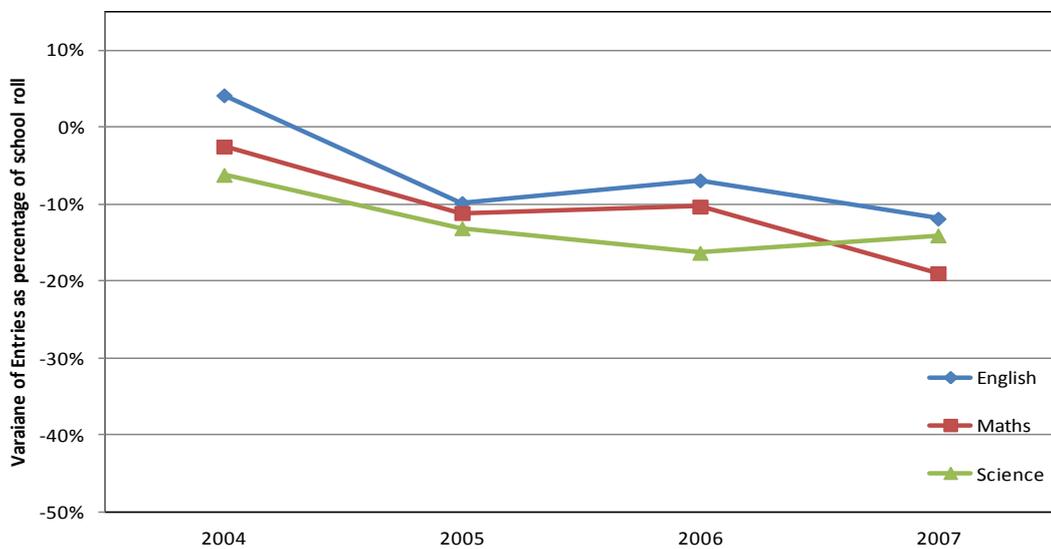


Figure 1: Percentages of Cluster Students Enrolled for Selected Subject Literacy Achievement Standards as Variance from National Percentages

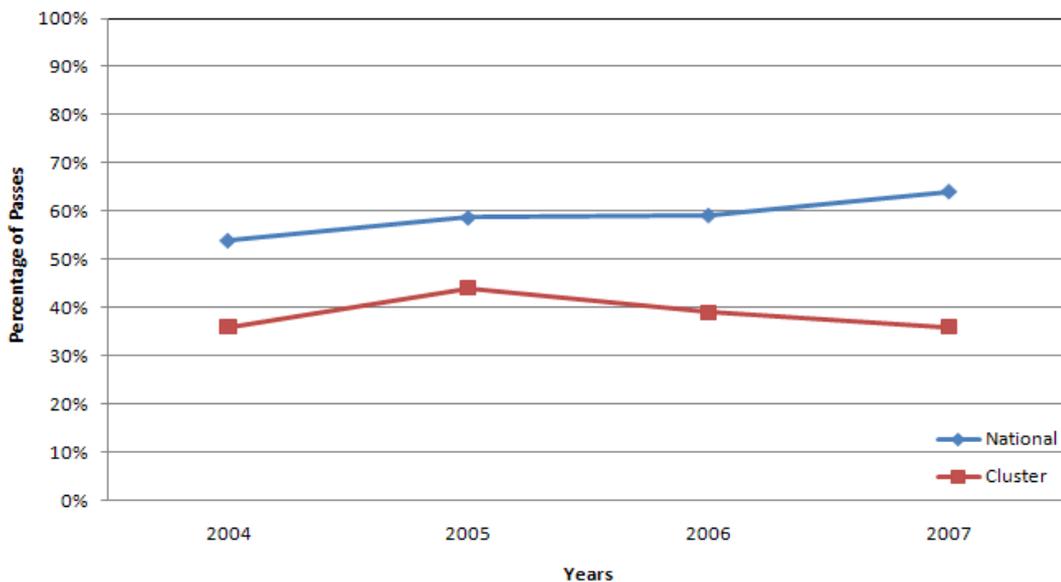


Figure 2: Percentage of Passes in Achievement Standard 90057: “Read and show understanding of unfamiliar texts”

We examined one core literacy standard in the NCEA level 1 exam (see Figure 3). This is an externally examined standard and is an indicator for effective literacy levels, the closest to the sorts of skills assessed on STAR and aTTle. The pattern from 2004 to 2007 is relatively stable, as is the national pattern. Students in the region who sat Achievement Standard 90057 were consistently less successful than students nationally (about 20-30% below). This suggests that despite higher reading comprehension at Year 8 and the beginning of Year 9, students found an equivalent task more challenging than would be predicted, implicating specific literacy instruction.

Further testing this hypothesis, lessons were analysed in terms of the frequency and quality of instruction in reading, writing, and vocabulary. For example, the following utterance by a Year 9 Social Science teacher was coded as an example of explicit reading instruction:

Remember if you are reading a news article, a lot of the time what is at the start of it and what is at the end of it are the really important bits... because they have to write something very, very briefly, they are not writing a book, they are writing a newspaper article so it's only going to be very, very short and has to get straight to the point and the beginning of it is very, very concise. (Social Science teacher, classroom observation, 15 October, 2008)

The analysis showed what we judged to be less than optimal rates of effective literacy instruction. Only 26% (8/31) lessons had at least one instance of explicit reading instruction, 23% (7/31) had at least one instance of explicit writing instruction, and 68% (21/31) had at least one instance of vocabulary instruction. Of a total of 78 utterances about vocabulary, 72% related to subject specialist academic vocabulary, 24% to general vocabulary, and 4% to general academic vocabulary. While vocabulary instruction was evident in more lessons (68%), most vocabulary instruction was restricted to subject-specific terminology, rather than general academic or low frequency items, and to receptive, rather than productive vocabulary. We retained this hypothesis on the basis of this evidence.

A related dimension concerns the nature of teacher-student relationships and the extent to which they are academically focused, which might contribute to less than desirable rates of student achievement (e.g., Bishop et al., 2007). Qualitative analysis indicated that the general classroom socialisation dynamic did not sufficiently focus on high-level academic engagement. The students' responses also indicated this. The prevailing climate was characterised by teacher-student relationships that were generally familiar, positive and respectful on an interpersonal basis but appeared to be less academically challenging for students. For example, while there were numerous exchanges that demonstrated that teachers knew and respected their students as individuals, there were relatively fewer teacher requests for students to improve the output or quality of their work. There did not appear to be a differentiation in terms of ethnicity. This form of the instructional hypothesis was retained.

General Discussion

An achievement problem in a group of secondary schools was identified. It was a mismatch between reading achievement at Year 8 and subsequent levels of achievement three years later. This mismatch was examined for two purposes. One was to describe the results of the first phase in the application of a school change model, the LSM (McNaughton et al., in press), to this problem as the basis for a localised intervention. The first phase is a process of profiling a problem. Using the results, the second purpose was to add to knowledge of the developmental transition between primary and secondary schooling.

The explanations for the achievement patterns in this region were largely discounted. Being able to discount these was important also because of the possibility that they had contributed to beliefs about limits on effectiveness due to phenomena largely outside of one's control and hence limited collective school efficacy (Bandura, 1995). In previous interventions discussing beliefs about causes, especially those that attribute failure to deficits in the students, their families or their communities, are important to building collective efficacy (e.g., Bishop et al., 2007; Lai et al., 2009).

Research on the shift from primary to secondary suggests a transition effect that mostly happens early in secondary school (Galton et al., 2009; Hogden et al., 2009). The effect has been tied to motivational and attitudinal outcomes of a shift in pedagogy from primary to secondary. We didn't examine the relationship between the pedagogies at primary and at secondary but used two sources of evidence to check this hypothesis and its corollary. We examined the timing of the possible impact of the transition on students' achievement and we also probed their attitudes to school. The effects were not noticeable immediately (about four weeks into school) but were increasingly apparent over two years which implicated aspects of the instructional environment. Other studies have also plotted transition effects both in New Zealand (Hogden et al., 2009) and elsewhere (Braund, 2008; Galton, 2009). These studies focus on the motivational characteristics of students but tie the changes in motivation to contrasts in pedagogy. The present study provides some support for this association but also suggests for these students the effect was exaggerated by specific features of pedagogy.

Several aspects of the hypothesis about instructional quality were retained after testing by the combined group of researchers and leaders. The general pedagogy of the classroom did not look very different to other secondary classrooms in New Zealand. This raised the question that even if, at a national level, there is some sort of contrast effect or increasingly powerful effect of a qualitatively different form of pedagogy from primary school, why would it be exaggerated in these secondary schools? Several data sources suggested a more subtle instructional effect relating on the one hand to the degree of academic focus (seen in relatively low expectations and cognitive demand) which were corroborated by student interviews, and on the other to specific content area literacy needs (Draper, 2008; Jetton & Alexander, 2004; McDonald & Thornley, 2005).

The implication of this profiling for the ongoing application of the LSM was the need for professional development targeted at effective instruction in content area literacy and changed expectations relating to academic success through the leadership in professional learning communities within the schools. The wider significance of these data is to demonstrate the need to contextualise school change or school intervention programmes. Without the process of profiling described here, ineffectual, inappropriately targeted or unsustainable interventions may occur (Lai et al., 2009).

This analysis deliberately 'contextualises' the problem and a planned limitation in the study is its limited generalisability. Generalising beyond these schools is not appropriate. However, the transition phenomenon established in the literature and other New Zealand research (Bishop et al., 2007) suggests that where there are transition effects, a fruitful focus would be those dimensions of secondary pedagogy that construct expectations, high challenge, and cognitive load. The study is limited in another way in that there are questions to do with how the processes identified in the secondary schools differentially affect students with different backgrounds and characteristics. While this was not the case in this study, others have found important differences in patterns (e.g., Hogden et al., 2009).

Acknowledgements

We warmly acknowledge the substantial contribution that different groups in the practice and learning community have made to the successful completion of the research project. We especially acknowledge the students, teachers, principals and literacy leaders at the schools with whom we worked for their time and contribution.

References

- Bandura, A. (1995). Exercise of personal and collective efficacy in changing societies. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 1-45). Cambridge, England: Cambridge University Press.
- Benner, A. D., & Graham, S. (2009). The transition to high school as a developmental process among multiethnic urban youth. *Child Development, 80*(3), 356-376.
- Biddulph, F., Biddulph, J., & Biddulph, C. (2003). *The complexity of community and family influences on children's achievement in New Zealand: Best evidence synthesis*. Wellington, New Zealand: Ministry of Education.
- Bishop, R., Berryman, M., Powell, A., & Teddy, L. (2007). *Te Kotahitanga: Improving the educational achievement of Māori students in mainstream education. Phase 2: Towards a whole school approach. Report to the Ministry of Education*. Wellington, New Zealand: Ministry of Education.
- Bishop, R., O'Sullivan, D., & Berryman, M. (2010). Evidence. In R. Bishop, D. O'Sullivan, & M. Berryman, *Scaling up education reform: Addressing the politics of disparity* (pp. 131-164). Wellington, New Zealand: NZCER Press.
- Bransford, J., Derry, S., Berliner, D., & Hammerness, K. (2005). Theories of learning and their roles in teaching. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 40-87). San Francisco, CA: Wiley.
- Braund, M. (2008). *Starting science ...again?* London, England: Sage.
- Callanan, M., Kinsella, R., Graham, J., Turczuk, O., & Finch, S., (2009). *Pupils with declining attainment at key stages 3 and 4: Profiles, experiences and impacts of underachievement and disengagement*. London, England: Department of Children, Schools and Families.
- Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing: When the 'have nots' gain but the 'haves' gain even more. *American Psychologist, 60*(2), 149-160.
- Draper, R. J. (2008). Redefining content-area literacy teacher education: Finding my voice through collaboration. *Harvard Educational Review, 78*(1), 60-83.
- Education Review Office. (2003). *Schooling on the West Coast*. Wellington, New Zealand: Author.
- Elley, W. (2001) *STAR supplementary tests of achievement in reading: Teachers' manual*. Wellington, New Zealand: New Zealand Council for Educational Research.
- Galton, M. (2009). Moving to secondary school: Initial encounters and their effects. *Perspectives on Education, 2*, 5-21.
- Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London, England: Routledge.
- Hattie, J. A. C., Brown, G. T. L., Keegan, P. J., MacKay, A. J., Irving, S. E., Patel, P., ... Campbell, A. R. T. (2004). *Assessment tools for teaching and learning (asTTle) version 4, 2005*. Wellington, NZ: University of Auckland / Ministry of Education / Learning Media.
- Hogden, E., Ferral, H., & Dingle, R. (2009). *Technical report with growing independence: Competent learners @14*. Wellington, New Zealand: Ministry of Education. Retrieved from http://www.educationcounts.govt.nz/__data/assets/pdf_file/0017/7721/cc-technical-independence14.pdf
- Hong, S., & Ho, H-Z. (2005). Direct and indirect longitudinal effects of parental involvement on student achievement: Second-order latent growth modeling across ethnic groups. *Journal of Educational Psychology, 97*(1), 32-42.
- Jetton, T. L., & Alexander, P. A. (2004). Domains, teaching, and literacy. In T. L. Jetton & J. A. Dole (Eds.), *Adolescent literacy research and practice* (pp. 15-36). New York, NY: Guilford Press.
- Jia, Y., Way, N., Ling, G., Yoshikawa, H., Chen, X., Hughes, D. ... Lu, Z. (2009). The influence of students' perceptions of school climate on socioemotional and academic adjustment: A comparison of Chinese and American adolescents. *Child Development, 80*(5), 1514-1530.

- Lai, M. K. (2009). *West Coast secondary project: Milestone 2 report*. Auckland, New Zealand: Auckland UniServices.
- Lai, M. K., McNaughton, S., Amituanai-Toloa, M., Turner, R., & Hsiao, S. (2009). Sustained acceleration of achievement in reading comprehension: The New Zealand experience. *Reading Research Quarterly*, 44(1), 30-56.
- Levin, B. (2008). *How to change 5000 schools: A practical and positive approach for leading change at every level*. Cambridge, MA: Harvard Education Press.
- McDonald, T., & Thornley, C. (2005). Literacy teaching and learning during the secondary years: Establishing a pathway for success to NCEA and beyond. *set: Research Information for Teachers*, 2, 9-14.
- McGee, C., Ward, R., Gibbons, J., & Harlow, A. (2004). *Transition to secondary school: A literature review*. Hamilton, New Zealand: Waikato Institute for Research in Learning & Curriculum, University of Waikato.
- McKinley, E., Madjar, I., van der Merwe, A., Smith, S., Sutherland, S., & Yuan, J. (2009). *Targets and talk: Evaluation of an evidence based academic counselling programme*. Auckland, New Zealand: Starpath Project, The University of Auckland.
- McNaughton, S. (2002). *Meeting of minds*. Wellington, New Zealand: Learning Media.
- McNaughton, S. (2011). *Designing better schools for culturally and linguistically diverse children: A science of performance model for research*. New York, NY: Routledge.
- McNaughton, S., Lai, M. K., & Hsiao, S. (in press). Critical foci for data use: Raising literacy achievement levels across a cluster of schools. *School Effectiveness and School Improvement*.
- Morris, A. K., & Hiebert, J. (2011). Creating shared instructional products: An alternative approach to improving teaching. *Educational Researcher*, 40(1), 5-14.
- O'Brien, P., & Lai, M. K. (2011, September). *Year 10 reading comprehension: Year 10 student perspectives on achievement*. Presentation to the Symposium on Assessment and Learner Outcomes, Victoria University of Wellington, New Zealand.
- Raphael, T. E., Au, I. K. H., & Goldman, S. R. (2009, April). *A developmental model of the standards-based change process: A case study of school literacy reform*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, California.
- Rowan, B., Correnti, R., Miller, R., & Camburn, E. M. (2009). School improvement by design: Lessons from a study of comprehensive school reform programs. In G. Sykes, B. L. Schneider & D. N. Plank (Eds.), *Handbook of education policy research*. Washington, DC: American Educational Research Association, Routledge.
- Wang, J. H., & Guthrie, J. T. (2004). Modeling the effects of intrinsic motivation, extrinsic motivation, amount of reading, and past reading achievement on text comprehension between US and Chinese students. *Reading Research Quarterly*, 39, 162-186.
- Wang, Q., & Pomerantz, E. M. (2009). The motivational landscape of early adolescence in the US and China: A longitudinal study. *Child Development*, 80, 1272-1287.

New Zealand's First Tertiary High School: Students' Perspectives of Various Aspects of Their Education and Experiences in School

Colleen Young

Manukau Institute of Technology and the University of Melbourne

Email: colleen.young@manukau.ac.nz

Abstract

Given that economists predict that students in the future will need not just a senior secondary school credential but a postsecondary one to enter the workforce, it is imperative that alternative forms of schooling are sought to re-engage students who are failing within the current senior schooling infrastructure. This study forms part of a longitudinal study aiming to evaluate the determinants of student success at New Zealand's first Tertiary High School (THS) within the first two years of operation. The THS commenced operation in 2010 with 43 Year 11 students in collaboration with 12 schools in South Auckland. This paper reports on the students' perceptions of their THS experience in comparison to their perceptions of their traditional schooling experience in their first year. The study used a qualitative method that incorporated both a narrative approach and the use of "program logic" (an evaluation framework) to gain a deeper understanding of seven, randomly selected, Year 11 students' experiences and views of the new form of schooling. Findings show that the students perceived the change in the schooling environment assisted them to be well on the way to achieving their NCEA Level 1 by year end.

Keywords: Tertiary High School, program logic, transition, disengagement

Introduction

The purpose of this paper is to demonstrate through students' logic and narrative how they viewed their experiences within the traditional high school setting in contrast with their views on attending New Zealand's first Tertiary High School (THS) in its first year of operation. The THS is located in the South Auckland region of New Zealand. The paper begins with a brief discussion on student failure in the traditional school system followed by a review of several alternative forms of schooling and how they affect student outcomes. The next sections outline the research methods, the findings with a brief discussion, conclusion, and ideas for future research that could strengthen and complement the findings in this paper.

Student Disengagement

Research shows that the "one-size-fits-all" traditional education system is no longer working for underrepresented minority students and this is causing increased student disengagement and dropout rates in high schools (Kazis 2004; Kirst, 2004). There are several explanations for this such as: the type of curriculum offered to senior students (Stern & Stearns, 2008); a lack of academic and social support mechanisms for students who need more guidance (Hooker & Brand, 2009); student 'knowledge gaps' in terms of their future career pathways (Stern & Stearns, 2008); a lack of "college knowledge" or preparedness for some senior students to make a smooth transition to tertiary education or the workplace (Conklin & Sanford, 2007), and a lack of students' understanding of the need to set goals in which their activities align with working towards achieving their desired outcomes (Biggs, 2003).

The Extent of the Issue

An increase in student disengagement and drop-out rates has not only ramifications for individual students but it also impacts on society and economic productivity. Cohen (1996) has estimated the dollar value of “saving” an at-risk youth at US\$1.5 - \$2.0 million, because costs such as dropping out of society, or having the potential for using drugs are extra costs to the justice system. Similarly, in the United Kingdom, Hayes (2007) argued that there were about 1.3 million not in education, employment or training leading to an annual cost to the United Kingdom of £3.65 billion.

Moreover, student disengagement also has social risks. Rumberger (2001) argued that students who drop out are less likely to find work and more likely to live in poor conditions, become unhealthy or even be incarcerated than their peer group who graduate from high school. Given that the number of students dropping out of the education system is so high, there is continuing demand to find new ways to re-engage these students.

One Solution: Multiple Pathways

A solution often promoted to re-engage students in learning is the concept of Multiple Pathways (Carnevale & Desrochers, 2002; Oakes & Saunders, 2008; Rogers, Kahne, & Middaugh, 2007; Schwartz, 2004; Steinberg & Almeida, 2007). Multiple Pathways involves creating individual and flexible learning pathways for all students (Schwartz, 2004).

Below is a brief critique of several interventions that demonstrate varying levels of success in improving student outcomes in the United States, Australia, United Kingdom and New Zealand. They are: Early College High Schools, Careers Academies, Talent Development High Schools, Dual Enrolment, Senior Colleges, Australian Technical Colleges and Trades Training Centres, Vocational Education Training in Schools, Academies, Gateway, Secondary-Tertiary Alignment Resource, Youth Guarantee Programme and Trades Academies and Service Academies.

Early College High Schools

In 2002, Early College High Schools (ECHSs) emerged in the United States supported by the Bill and Melinda Gates Foundation. There are now over 240 ECHSs across 28 states and the District of Columbia serving more than 75,000 students. The key goal was to assist low socioeconomic, under-represented youth, first-generation college goers, to concurrently gain college credits, while still studying at high school, at no cost to the students. According to Hoffman and Vargas (2005), the ECHS model was established to “test the hypothesis that, with proper support, students at risk of not gaining postsecondary credentials benefit from taking college-level courses and earning college credit while in high school” (p. 3). An evaluation using a logic model by the American Institutes for Research and SRI International over the past seven years has shown that ECHS students, on average, are performing better than students in the surrounding school districts in attendance and achievement (Berger et al., 2008).

Career Academies

Career Academies, a United States intervention, were established in the 1970s and there are now over 2,500 operating nationwide. Career Academies target students from grades 10-12, combining both academic and technical curricula that incorporate a career theme (Kemple, 2008). For the past four decades, Career Academies have prepared students for successful transitions from secondary to postsecondary level and into the workforce (Bragg, Kim, & Rubin, 2005; Conley, 2010; Hooker & Brand, 2009; Kemple, 2008). Beginning in 1993, the Manpower Demonstration Research Corporation, a non-profit, non-partisan social and education policy research organisation, has conducted evaluation research on the Career Academy model (Kemple, 2008). Findings show evidence that Career Academy students benefit from being better prepared for the workforce and

employment outcomes – especially for ‘at-risk’ young men – and they experienced more successful transitions from school to work than their counterparts who had not participated in the Career Academy model. Stern, Raby, and Dayton (1992) found that Career Academy students achieved better grades and had improved attendance compared to non-Career Academy students. In addition, former Career Academy students demonstrated positive effects in terms of their employment and incomes, their family relationships and independence eight years after they had left when they were compared to non-Career Academy students (Hooker & Brand, 2009; Kemple, 2008).

Talent Development High Schools

Talent Development High Schools (TDHSs) were developed by Johns Hopkins University and now operate in 15 states of the US. The key concept of TDHSs is to promote smaller schools within a failing larger school and to focus on improving achievement grades in English and mathematics. In conjunction with a smaller school environment, the school builds lasting relationships with the wider community and endeavours to engage the whole family with the school. Research using a quasi-experimental design over a four-year period found that 68% of students from TDHSs progressed to Grade 10 compared to only 60% in the control group (Tyler & Lofstrom, 2009).

Dual Enrolment

Dual enrolment integrates senior secondary school courses with postsecondary learning programmes. In the United States, over 813,000 students per annum dual enrol for courses in over 38 states (Kleiner & Lewis, 2005). Evidence shows that participating dual enrolment students were more likely to graduate from high school, enrol in college, and gain higher post-secondary grade-point averages than the non-participating students. College courses are sometimes free or discounted, saving money for families that may not be in a position to send their child to college. Students can undertake their learning either at their local high school, on a college campus or study through a distance learning provider. Dual enrolment students, when compared with a similar non-dual enrolment group, were also more likely to enrol in four-year institutions which could mean that students had improved confidence from their early postsecondary experience leading also to higher educational aspirations (Hughes, 2010).

Senior Colleges

In Australia, one solution to ease the secondary-postsecondary transition process for students who are likely to drop out in the traditional schooling environment has been the creation of senior colleges. Senior colleges replaced existing senior high schools in the early 1980s with the first two commencing operations in 1983 (Fraser, Williamson, & Tobin, 1988). Senior colleges accept all senior high school students and some schools also accept re-entry adult students. The curriculum offers a range of options from university oriented courses to Vocational Educational Training (VET) programmes. Research claimed that the students in senior colleges: had access to a wide range of subjects including VET programmes; preferred the adult learning environment; and improved their confidence which in turn eased their transition to the workplace or further study (Polesel, 2002). In addition, the collaboration between the senior school sector and the tertiary sector allowed for sharing of resources, teaching and facilities, and the demand for places grew over time making student selection more competitive (Polesel, 2002).

Australian Technical Colleges, Trades Training Centres and the Apprenticeship programmes

In 2005, the Australian Federal Government funded the Australian Technical Colleges (ATCs) to deliver vocational educational courses to senior students in Year 11 and Year 12. By 2009 the ATCs had expanded to 24 sites but the government closed the programme in December 2009. In response to the closures, the 24 ATCs were integrated into the Trade Training Centres (TTCs) in Schools Program and into Australian

Apprenticeships. TTCs in Schools work with local industries and training partners to support Indigenous students and disadvantaged students. The main objective is to improve educational outcomes for these students in order to ease the transition from school to further education or build and improve upon pathways that lead to work in the local community (www.tradetrainingcentres.deewr.gov.au).

In 2008, funding of \$2.5 billion was invested into TTCs and this was the first time in Australia that all senior secondary students were able to access vocational training at school. By December 2011, the TTCs in Schools programme was operating in 1060 schools which is close to 40% of the eligible schools, 146 TTCs had been opened, and a further 88 TTCs were under construction. Initial findings show promise for students who have enrolled in the TTCs programme. By the end of 2010 the TTCs programme had enrolled more than 1,600 students in 60 schools with around 90% of the students achieving an accredited trade qualification, or continuing to be engaged in further training.

In addition, apprenticeships are available to people returning to the workforce or wishing to change their career pathway, or for senior high school students in Year 11 or Year 12. Careers advisers and teachers in schools assist the students to find the pathway that suits them as there are more than 500 occupational choices.

Vocational Education and Training in Schools

The VET programme in Australia aims to support senior Year 11 and Year 12 students by integrating both the traditional academic studies in vocational courses and providing planning and assistance to students to pursue a career pathway in the vocational area (www.deewr.gov.au/schooling/careersandtransitions/vocationallearning).

Gateway

A New Zealand initiative, established in 2001, the Gateway programme utilises a collaborative approach working with employers, schools and students. The programme allows students to benefit from learning in a dual environment, both in school and in their local workplace, and it operates within 24 schools reaching 1,000 students and over 200 employers. It is aimed to re-engage “at-risk” youth studying at years 11-13+ who are in school. Some examples of student placements are in hairdressing, media, furniture and joinery, and tourism. During the first three years, a pilot evaluation commissioned by the Tertiary Education Commission (2003) identified several student benefits such as improved educational achievement, increased self-confidence, and more employment opportunities.

Secondary-Tertiary Alignment Resource

Introduced in 1996, the Secondary-Tertiary Alignment Resource (STAR) programme was established to assist students’ transition from school to postsecondary education or into the workplace. It allows schools to “purchase non-conventional tertiary level courses leading to credits on the National Qualifications Framework” (Vaughan & Kenneally, 2003, p. 1). The target audience for the STAR programme is all high school students. For example, short taster courses can be undertaken by any students in both Year 9 and Year 10. At a senior secondary school level – Year 11, 12 and 13 – students can enrol for full courses. In 2003, findings demonstrated a mixed reaction to using a ‘student focused’ approach as two-thirds of the participating schools used the programme to keep students enrolled where others assisted the students into specific employment pathways or further training (Vaughan & Kenneally, 2003).

Youth Guarantee Programme

A government funded initiative, the Youth Guarantee Scheme was launched in New Zealand in January 2010 with the goal of targeting 16-17 year-olds who feel they are not the right ‘fit’ for the traditional school system but may like to try a range of tertiary courses at no cost to them. The Youth Guarantee programme has been established to assist in bridging the gap between secondary and postsecondary education. Students receive not

just technical training but support in learning, cultural, and emotional areas. For example, students are provided with career planning or attending orientation days on the tertiary campus to get to know some of the other students. Initially the programme was available for 2,000 students; however, it is too early to ascertain whether the Youth Guarantee programme is improving student outcomes for underrepresented, disengaged youth.

Trades Academies

According to the Ministry of Education website, “Trades Academies aim to get young New Zealanders engaged in education, and equip tomorrow’s workforce with relevant skills by linking with the wider industry training system ... Partnerships are formed between schools, tertiary institutions, industry training organisations and employers” (p. 1). Students are required to complete a two-year programme to the standard accepted in a trade that they have work-trialled. Each student can gain credits in the National Certificate of Educational achievement (NCEA) while simultaneously earning a tertiary qualification. Currently, the aim is to create five Trades Academies during the next three years. There has been no formal evaluation of the effect Trades Academies are having on student outcomes.

Service Academies

Service Academies are similar to Trades Academies in schools but are military focused. Established in 2009, there are now 18 Service Academies operating in New Zealand. Service Academies target senior high school students who are ‘at risk’ of dropping out of school. Approximately 20 students attend each Service Academy for a year. The programme incorporates a combination of courses by the New Zealand Defence Force, leadership activities, and NCEA Level mathematics and English courses. It is too early to tell whether the programme is improving student outcomes. According to the website, the purpose of Service Academies is to:

- encourage students to stay engaged in learning by providing a motivating and disciplined programme
- help students to gain improved qualifications and help them prepare to move successfully into the workforce or further education and training.

Tertiary High School

The THS evolved as a response to the issue of student disengagement within a traditional school setting. Established in February 2010, 46 Year 11 students became dual enrolled at Manukau Institute of Technology (MIT) and their previous school. Utilising a collaborative approach, students were selected from the surrounding secondary schools on a recommendation from their counsellor, principal, dean or member of the teaching staff and identified as ‘likely to fail’ within the existing school framework. The initial student intake came from 12 high schools in the South Auckland region (Middleton, 2010).

The key goals for the THS are threefold. The first aim is to integrate the final secondary years of education with the first years of tertiary education and training, at no expense to the student; the second is to promote students’ learning at their own pace for the acquisition of qualifications at both secondary and tertiary levels; and third is to blend academic learning with vocational courses for students in order to create a career pathway that is challenging, rigorous, and relevant to them (Middleton, 2009).

The THS program contains a flexible series of courses at MIT called Preparation for Trades and Professions (PTPs) which include courses such as: automotive, fabrication, early childhood education, and catering and hospitality, and combines these “tasters” with the relevant numeracy, literacy and technological skills required for any future career path a student may choose. In addition, the THS engages the student in an individual learning programme which is monitored regularly by their form teacher (lecturer) in terms of credits achieved and relevance to their chosen career path. Social and academic support is

offered to students on a “needs basis” but the concepts of mentoring, goal setting, and reflecting on their results is incorporated into the THS programme (Middleton, 2008).

National Certificate of Educational Achievement

Introduced in 2002, the NCEA is New Zealand’s key qualification for high school students and is recognised by educational institutions both nationally and overseas. Each senior year at secondary school usually correlates with an NCEA level of achievement – NCEA Level 1 typically is completed in Year 11, NCEA Level 2 in Year 12 and NCEA Level 3 in Year 13. It is a complex qualification system as it endeavours to cater for students with a wide range of knowledge, skills and interests. For example, on a learning continuum, students can progress to occupational training ranging from medicine to hairdressing or join the workplace immediately after school (Madjar, McKinley, Deynzer, & Van Der Merwe, 2010). Each student aims to gain a certain number of credits at each level (e.g., NCEA Level 1 requires 80 credits to pass). These credits are obtained via internal and external assessments and can be passed at three different achievement levels: Achieved, Merit or Excellence. If Not Achieved, the student fails to be awarded credits.

Methods

The aim is to show the differences and/or the similarities between the students’ programme logics related to what they considered was the logic if they had remained at their traditional school compared to their logic in the THS. Semi-structured in-depth face-to-face interviews were conducted with seven randomly selected participants at the THS in 2010, followed by focus groups (Punch, 2005) with the same participants, in order to validate and triangulate the adopted logic models. To protect student confidentiality, pseudonyms have been used for each of the students. The small sample size precludes these results from being generalised. The findings include program logic models of the students’ logics for both forms of schooling combined with a narrative approach to gain a deeper understanding of the students’ views (Clandinin & Connelly, 2000).

Program Logic as a Theoretical Framework

Logic models are a form of evaluation that dates back to the 1970s (Bennett, 1976; Wholey, 1979). Program logic is a diagram that provides a theoretical framework for understanding an intervention and how its actions are intended to impact on short-term, medium-term and long-term goals. Program logic is flexible in nature and evolves over time. In its simplest form, program logic assesses inputs, processes, outputs and outcomes (Kellogg Foundation, 2008). According to Duignan (2009), who uses the term outcomes model instead of program logic, an outcomes model can give a rich representation of what is believed to cause what in response to an intervention. In essence, program logic depicts a theory of action that demonstrates cause and effect – if you “do” this, then “that” will happen. Program logic is not rigid in its form as there are many variations of logic models.

Program logic is not reality, but it can be used as a story-board to view what resourcing is required and what steps need to be taken for the stakeholders to achieve their goals. In doing so, the program logic model can become an integral part of the theory of change. An alternative “improved” program logic model can be developed and this cycle can be repeated over time.

There are several benefits in utilising a program logic model. First, acknowledging and incorporating the stakeholders’ views is a powerful way to gain a shared understanding and program vision (Taylor-Powell & Henert, 2008). Secondly, program logic is visual and stakeholders can immediately view the program inputs, process, and short and long-term outcomes on the one page. Finally, using program logic enables the stakeholders to prioritise the resources combined with viewing cause and effect from each step (Duignan, 2009).

In contrast, there are several limitations to the use of program logic. First, program logic does not represent reality but is a tool to visualise a program. The focus is on intended outcomes; consequently, program logic may overlook unintended outcomes and may simplify the complex nature of causal attribution where many factors influence process and outcomes. Furthermore, it may not address whether the stakeholders are doing the right thing (Taylor-Powell & Henert, 2008).

The value in using a program logic model is that it provides an immediate “snapshot” of inputs (resources including human resources), activities (programs implemented) and short-term and long-term outcomes (including achievement and other outcomes such as behavioural or community engagement) (Chen, 1990; Duignan, 2009). In other words, program logic offers a framework and stakeholders can ask the question – Have the resources available enabled the right actions to achieve the intended outcomes? If not, what resources would be required? What activities would need to be changed? Program logic, or a theory-of-action, can assist stakeholders to engage in conversation to improve a project or program.

Prior to interviewing the students, a group of experienced educators built, and agreed upon, their version of the program logic model. These stakeholders were selected as people highly involved in the project at the implementation stage. They ranged from the project instigator, THS principal, programme initiators, current teachers at the THS and MIT, and principals and careers advisors from schools in the surrounding area. In this study, during the in-depth interviews, the students were asked questions in relation to the logic model and were asked to compare their experiences in both the THS and their traditional school settings.

During the semi-structured interviews (May, 2001; Punch, 2005), the students discussed their past schooling experience and contrasted this experience with their new form of schooling at the THS. Prior to commencing the study, ethics approval was granted by the University of Auckland and MIT.

Participants

Below is an overview of the seven student participants outlining the reason why they changed schools and the initial goals they set for Year 1 at the THS:

Trent changed schools after a discussion with the Head of House at his previous school. He found his last school boring and hated doing normal school work so he never turned up to class. Trent would like to be an electrician.

Angela found out about the THS through her school counsellor and the deputy principal at her previous school. She was a major truant and thought that the teachers had no time for her. Angela has set her initial goal to pass NCEA Level 1.

John was reluctant to change schools even though he was underachieving at his previous school. His mother thought that trying a new form of schooling at the THS may improve his learning outcomes as he had a major truancy problem at his last school. John’s long-term goal is to be a mechanic.

Jody ‘wagged a lot’ at her last school and heard through other students already attending the THS that it was stricter than her previous school. She thought the THS learning environment would suit her better so she started at the THS in week three. Jody’s first two goals she set for herself was to attend school regularly and improve her learning outcomes.

Carol had relationship problems with others where she felt she was being bullied at her previous school. The school counsellor suggested to her to enrol at the THS. Carol soon made friends at her new school and she seeks to obtain 40 NCEA credits by September. Her long-term goal is to become a business person.

Steven enjoyed the elective subjects at his previous school such as automotive, fabrication and photography. The rest of the day was spent in the bush area behind the school. Steven made his own decision to change schools and viewed moving to the THS as making a fresh start. Steven wants to join the Air Force when he turns 18.

Julie struggled with school rules such as having to get a note prior to leaving the school grounds at lunch time or being assigned a certain seat in class. She heard about the THS and decided to attend an open day with her friend. Julie initially wanted to achieve NCEA Level 1 and long-term she wants to be an early childhood teacher.

Results

As illustrated in Figure 1, the program logic theory-of-action model depicts the students' views of the inputs, activities, and short and long-term outcomes at the THS. By contrast, Figure 2 demonstrates the program logic theory-of-action model of the students' perceptions of the inputs, activities, and short and long-term outcomes they may have experienced at their traditional school, had they remained there.

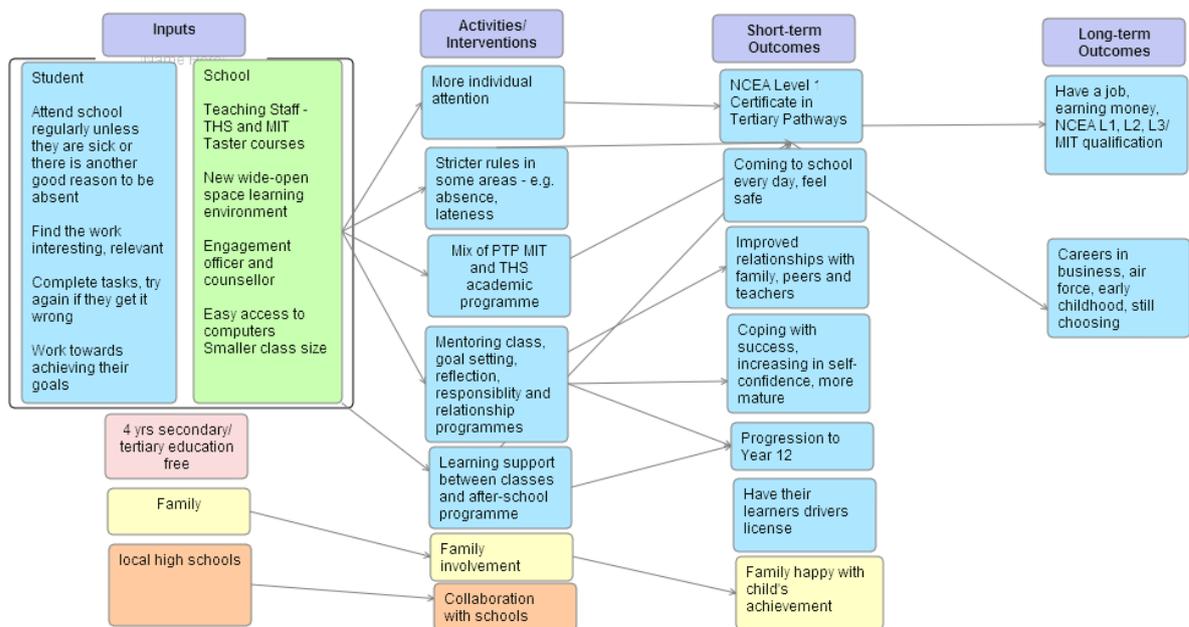


Figure 1: Program Logic Model: Students' Perspective – Tertiary High School, August 2010

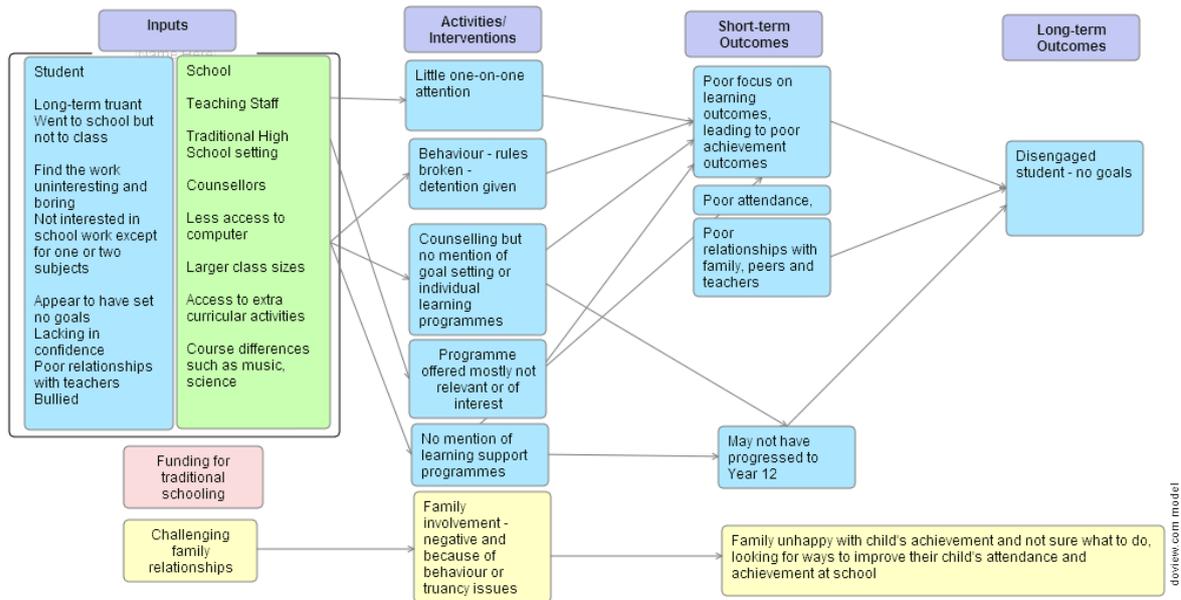


Figure 2: Program Logic Model: Students' Perspective – Traditional High School, August, 2010

Both program logic models begin with the inputs required such as the human resources and the capital resources for the school to operate. Inputs also included the students' perceptions of their own characteristics within both schooling environments. Finally, the external inputs are separated out such as the student's family and the wider community.

Column two of each program logic model depicts the specific activities or interventions according to the students' perceptions that were implemented in each school setting. These activities included the curriculum, academic and social support systems, family involvement and the wider community support.

The last two areas on the program logic diagram are the short and long-term outcomes. Short-term outcomes are goals that could be achieved within the first year. These included specific goals such as attending school regularly, academic achievement in Year 11, and progression to Year 12. Long-term outcomes demonstrate the students' perceptions of where they think they will end up within the next five years.

This section is divided into three parts – inputs, activities and short and long-term outcomes – to discuss the similarities and differences between the models. Student narrative is incorporated to further explain the students' progress in their first year.

Inputs

Whilst students claimed they attended school in both settings, the majority of them did not attend class regularly at their traditional school. This could indicate that a student simply turning up to school is not the only determinant of what constitutes student engagement. Figure 1 indicates that the students claimed that they set clear goals at the beginning of the year and that they attended the THS regularly unless they had a valid reason to be absent. Evidence suggests that the students were more engaged in learning and saw a purpose in their attendance once they understood their individual learning plan at the THS.

Steven: I came to MIT with a different frame of mind and also it was a better learning environment. I think cause in my last school there was just too many rules and I didn't like it. ... The place [THS] is loaded out with computers. The classes are less big as well.

In addition, the students set more specific goals which they felt were achievable. They felt that there was a wider range of courses on offer at the THS. They liked the wide-open spaced learning environment, smaller class sizes and easy access to computers, the engagement officer and guidance counselling. The students were also aware that the next four years of secondary / tertiary education for them was free while they were enrolled at the THS.

Jody: The THS program gives us a head start instead of going to school and finishing and then starting a [tertiary] course. Here we can start a [tertiary] course while we are still [school] students.

Trent: More hands-on work [at the THS].

Carol: I am wanting to achieve all my assessments in all my classrooms. My goal for this term is to come to school every day.

Angela: I just focus on my work and try to get my credits [at the THS].

Jody: For four years it is free for us for whatever course we want to do.

In contrast, in Figure 2 depicting the students' theory-of-action at their traditional school, they claimed that they found the work uninteresting and boring except for more practical subjects like automotive or photography. In addition, the students commented on their lack of any clear goals or any sense of urgency to attend school so they had a major truancy problem. Furthermore, they commented on the larger class sizes, less access to computers, and course differences such as music and science which are two subjects not offered to the students at the THS. In essence, the students' responses could be interpreted as their seeing no real purpose in what they were doing and where they were heading and therefore they slowly became disengaged.

Julie: At my last school I was quite bad – I didn't like the teachers or anything. I wasn't interested.

John: I had a lot of attendance issues at my old school 'cause I was kind of not very interested in going to school. ... Sometimes I'd have a full week off and sometimes I'd miss two days a week and three days a week.

Carol: I found it more difficult to learn there [at my previous school]. I had a bit of trouble with girls. Like having rumours about me and upsetting me a bit. Sort of losing confidence in going to school. ... I was getting bullied.

Activities

Figure 1 shows that, according to the students, they believed that THS staff implemented several different activities and interventions in order to achieve the students' expected short and long-term outcomes. First, students felt that they received more one-on-one attention at the THS. Second, they perceived that the rules within the THS were different – in some respects they were less strict as they could leave the premises at lunch-time and breaks as long as they returned on time for class, but in other respects the students felt that the THS was stricter. For example, the rules for attendance, lateness, and absence appeared to be much stricter. When a student turned up late for class the time had to be made up by the student at lunch time or after school. Teacher expectations appeared to be different. In addition, the students claimed that the mix of courses offered between MIT and the THS was different. The students also claimed that they received more mentoring, guidance in goal setting, reflection on their progress, and academic and social support at the THS in comparison to their previous schools. Finally, the students felt that the family (once they realised their son or daughter or sibling was making progress at school) was more involved at the THS than they were at their previous school from a positive, not negative, perspective.

- Steven: Well [at the THS, the teachers] don't just stand in front of the class. If you talk to them, they come to you and talk to you one on one. You don't have to tell the whole classroom what you want to know.
- Carol: I've still got to pass my other class – my other class last term. That is why we are doing mentoring.
- Carol: The teacher really helps us here. It is really good 'cause they push us to get our credits. They really help us.
- Julie: After school – we finish school at 2.30 but we don't. We go home at 3.00-3.15. They let us stay here if we want any help with anything or if we haven't done an assessment. We have a chance to catch up. Like any assessment.
- Steven: In our old school we would have to sit exams in strict conditions but here [at the THS] if you miss out on your credits you can re-sit them ... It's way simpler – you don't feel as pressured as you do when doing exams.
- Carol: I miss – sometimes at high school - you can get away with stuff too. But - not here - [you] can't get away with stuff here. It sort of makes you more mature.

In comparison, Figure 2 depicts the students' views of their previous schooling experience were that they received little one-on-one attention from the teaching staff and that if they were poorly behaved a detention would be given. There was no mention by the students of goal setting or being given individual learning programmes. Usual student counselling was provided for students which assisted some of them in changing to the THS. However, family only became involved in the school in a negative sense for student behaviour or truancy issues.

- Julie: The teachers at my last school] – They don't really open up. Like - um - they are kinda closed. They are like robots. Here you go. Here is the work. Do it.
- Jody: [At my previous school the teachers] give us instructions but we still don't get it. So then they just start yelling and we start yelling back and then the dramas.

In summary, according to the students' varying comments, the study findings suggest that the interventions implemented by the THS staff provided the students with a learning environment in which they felt they could succeed as compared to that in their traditional setting where they appeared to have no short or long-term goals and became disengaged.

Outcomes

Once the students understood what they needed to do to achieve their initial goals and they were given the guidance to complete individual tasks, their dialogue changed:

First, they set the goal to achieve NCEA Level 1.

- Carol: For this year – trying to get all my credits [for NCEA Level 1].
- Steven: Well I've got all the credits I can get so far. I only need 13 more to get NCEA Level 1."
- Trent: I will probably get into electrical engineering and get my NCEA Level 2.

In addition to the achievement outcome, the students recognised that they had to set more specific goals along the way such as attending school each day and improving their

self-confidence or building better relationships with their teachers, peers and family. This is reflected in comments made by Carol, Julie and Steven:

Carol: My goal for this term is to come to school every day.”

Julie: Well – I never really went to the dinner table last year. ... I never communicated with my family like I do now. ... I used to just grab dinner and go to my room. ... I've grown up ... you come here [to the THS and] they treat you as an adult and in your mind you are an adult. ... [In relation to my family]

Steven: My Dad supports it [the THS]. He reckons I wouldn't have got far at high school. Neither did he. He dropped out the first year he was there. ... So did Mum.

Several students had some idea of their long-term career path goals such as Steven, Carol and Julie when they commented:

Steven: When I found out about this course I wanted to do it so that I could get all my requirements so that I could join the Air Force when I was 18.

Carol: I want to be a business person. I want to run my own business.

Julie: [I want to do] early childhood education. ...Probably have a car because I would be [5 years] older then. Probably - success in NCEA. I think it would be Level 3 'cause you have to have Level 2 to get into ECE, so I have to stay here another year so I can go into that actual course.

Alternatively, Figure 2 represents the students' claims in terms of outcomes. They perceive that they may not have progressed to Year 12. This is reflected in their poor focus on learning outcomes, poor attendance and their interpretation of how they had poor relationships with their family, and peers and teachers at their previous school. In addition, according to the students, their families were not happy with their attendance, behaviour or achievement records and were not sure what to do about the problem. The end result, students' claimed, was that they set no goals and saw little relevance in attending school:

Jody: My Mum would like yell at me and tell me to go to class [at my previous school.] It went in one ear and out the other!

Angela: I had a lot of problems in my previous school. ... I hardly went to school. I didn't like school. ... I just stayed home or I would go with my friends somewhere.

Steven: Oh, well, I got kind of told off by my parents [when I didn't attend school] and also I found out if I just did nothing in school and then I would become nothing.

In short, the students' logic model presented in the THS in Figure 2 demonstrated that when they were provided guidance to set their own 'road map' of where they were heading in both the short and long term, their attitudes and behaviours shifted in order for them to achieve their goals. They felt they now had a purpose in going to school and began to take more ownership of their own life's journey.

Discussion

The results within the two logic models demonstrate the changes in student perceptions of their characteristics in the inputs section. This section discusses what the key determinants were that changed their behaviour through the THS teaching and learning activities.

First, the students immediately identified a reason to come to school in the THS as they had been shown how to set achievable short and long-term goals. Learning and achievement outcomes make more sense to a student when a student understands their short-term goals and the teaching and learning environment relates directly to these goals (Biggs, 2003).

Second, once the students saw a reason to come to class, their attendance record improved at the THS. Attendance alone was not the only determinant for student success; however, the improved attendance meant they were more likely to achieve their goals. According to Epstein and Sheldon (2002), when a student increases their class attendance record they are less likely to drop out and they are more likely to gain improved achievement outcomes than their counterparts who were absent more often.

Third, the integrated learning programme combining tertiary and secondary courses was a significant difference as the students claimed that the course provided more appropriate choices for them and they found the work both more interesting and more relevant than their classes at their previous school. These claims are in alignment with the ECHS Initiative Evaluation Conceptual Framework (2008) which included a blended curriculum incorporating academic rigour and relevant instruction for their students as one of the key determinants for achieving student success.

Finally, the rules for both forms of schooling differed and within the THS setting the rules assisted the students to gain an understanding of time – rules such as no bells at the THS, so when the students were late at the THS they were required to make up that time in breaks or after school with a teacher present to assist them if they needed help. This is vastly different from the students in their traditional school setting who claimed that they were given a detention for being late.

Conclusion

As illustrated in this paper, the value of utilising a logic model for this study was that it portrayed an immediate “snapshot” of inputs, activities and intended short-term and long-term outcomes according to the students’ logics for both the THS and the traditional forms of schooling. In some ways, the inputs were similar for both forms of schooling as both provided students with a place to learn and achieve academic outcomes; however, there were differences. This study suggests that the teaching and learning interventions initiated at the THS provided the students with an environment in which they felt that they could succeed. This contrasts with their traditional school environment where they believed they were failing. These varying responses from the students implied that the teachers at the THS were prepared to back them, give them a second chance, and work with them to create an individual learning path that made sense to them. According to this group of students, the THS teaching and learning interventions provided a new platform which gave them the right tools to believe they could achieve their desired goals, alongside encouraging them. In turn, this increased their confidence and motivation for them to achieve their short-term goals and look forward to obtaining their NCEA Level 1.

Future Research

As this study forms part of a longitudinal study, it only portrays the students' perspectives on both their past traditional schooling experience and the first year of their Tertiary High Schooling experience. Two further studies include the adult stakeholders' program logic model of the THS and a quantitative study researching progression, attendance rates and achievement outcomes for the THS students in Year 1 and Year 2 in contrast with a similar control group of students from high schools in the surrounding region. In addition, future research could focus on conducting interviews with former teachers of the THS students to further ascertain the extent to which these students have contributed to their own challenges in their past schools. To understand the full potential of the THS programme, future research showing how these students have done once they have left the education setting and entered the workplace would be useful in terms of replication of the THS in other regions both in New Zealand and abroad. To conclude, this study documented the responses of seven Year 11 THS students in the first year. The results clearly show the THS programme is very promising and that the findings appeared to confirm the logics of the assumptions made by the adult educational stakeholders who planned and implemented the THS programme and brought it to fruition. Completion of the two studies mentioned previously, combined with future research recommended in this paper will provide a clearer indication as to whether the THS programme can provide validation and support to be replicated both in New Zealand and overseas in the future.

References

- Australian Apprentice Centres. Retrieved from <http://www.australianapprenticeships.gov.au>
- Australian Technical Colleges. Retrieved from <http://www.deewr.gov.au/schooling/australian-technicalcolleges>
- Bennett, C. (1976). *Analysing impacts of extension programs, ESC-575*. Washington, DC: Department of Agriculture, Extension Service-US.
- Berger, A. R., Adelman, N., Cole, S., Cassidy, L., Darwin, M., Hall, C., et al. (2008). *2003-2007 Early college high school initiative evaluation: Emerging patterns and relationships*. Washington DC: American Institutes for Research.
- Biggs, J. B. (2003). *Teaching for quality learning at university* (2nd ed.). Buckingham: Open University Press/Society for Research into Higher Education.
- Bragg, D. D., Kim, E., & Rubin, M. B. (2005). *Academic pathways to college: Policies and practices of the fifty states to reach underserved students*. Paper presented at the Association for the Study of Higher Education, Philadelphia, PA.
- Bridgeland, J. M., Balfanz, R., Moore, L. A., & Friant, R. (2010). *Raising their voices: Engaging students, teachers, and parents to help end the school dropout epidemic*. Washington, DC: Civic Enterprises.
- Carnevale, A. P., & Desrochers, D. M. (2002). *The missing middle: Aligning education and the knowledge economy*. Washington, DC: US Department of Education, Office of Vocational and Adult Education.
- Chen, H. (1990). *Theory-driven evaluations*. Newbury Park, CA: Sage.
- Clandinin, D. J., & Connelly, F. M. (2000). *Narrative inquiry: Experience and story in qualitative research*. San Francisco: Jossey-Bass.
- Cohen, M. A. (1996). *The monetary value of saving a high-risk youth*. Nashville: Owen Graduate School of Management, Vanderbilt University.
- Conklin, K. D., & Sanford, S. (2007). A college-ready nation: An idea whose time has come. In N. Hoffman, J. Vargas, A. Venezia & M. Miller (Eds.), *Minding the gap: Why integrating high school with college makes sense and how to do it* (pp. 15-26). Cambridge: Harvard Education Press.
- Conley, D. (2010). *College and career ready: Helping all students succeed beyond high school*. San Francisco, CA: Jossey-Bass.
- Duignan, P. (2009). *Introduction to outcomes theory: Outcomes theory knowledge base Article No. 218*. Retrieved from <http://knol.google.com/k/paul-duignan-phd/introduction-to-outcomes-theory/2m7zd68aaz774/3>

- Early College High Schools*. Retrieved from <http://www.earlycolleges.org>
- Epstein, J. L., & Sheldon, S. B. (2002). Present and accounted for: Improving student attendance through family and community involvement. *Journal of Educational Research*, 95, 308-318.
- Fraser, B. J., Williamson, J. C., & Tobin, K. (1988). An evaluation of some senior colleges. *Journal of Educational Administration*, 26(3), 311-330.
- Hayes, J. (2007, April 24). The shameful growth of a generation of 'neets': John Hayes on 16-24 year olds not in education or training. *The Guardian*.
- Hoffman, N. (2003). College level credit in high school. *Change*, 35(4), 42-48.
- Hoffman, N., & Vargas, J. (2005). *Integrating grades 9 through 14: State policies to support and sustain early college high schools*. Boston, MA: Jobs for the Future.
- Hoffman, N., & Webb, M. (2009, June 11). Early college high school: Modest experiment or national movement? *Education Week*.
- Hoffman, N., Vargas, J., & Santos, J. (2009). New directions for dual enrollment: Creating stronger pathways from high school through college. *New Directions for Community Colleges*, 145, 43-58. doi: 10.1002/cc.354
- Hooker, S., & Brand, B. (2009). *Success at every step: How 23 programs support youth on the path to college and beyond*. American Youth Policy Forum. Available at <http://city.milwaukee.gov/ImageLibrary/Groups/cityDCD/LVP/pdf/SuccessAtEveryStep.pdf>
- Hughes, K., L. (2010). Dual enrollment: Postsecondary / secondary partnerships to prepare students. *Journal of College Science Teaching*, 39(6), 12-13.
- Kazis, R. (2004). Introduction. In R. Kazis, J. Vargas & N. Hoffman (Eds.), *Double the numbers: Increasing postsecondary credentials for underrepresented youth* (pp. 145-154). Cambridge: Harvard Education Press.
- Kemple, J. J., & Willner, C. J. (2008). *Career academies: Long-term impacts on labor market outcomes, educational attainment, and transitions to adulthood*. New York, NY: Manpower Demonstration Research Corporation.
- Kirst, M. W. (2004). Using a K-12 assessment for college placement. In R. Kazis, J. Vargas & N. Hoffman (Eds.), *Double the numbers: Increasing postsecondary credentials for underrepresented youth* (pp. 145-154). Cambridge: Harvard Education Press.
- Kleiner, B., & Lewis, L. (2005). *Dual enrollment of high school students at postsecondary institutions: 2002-03 (NCES 2005-008)*. Washington, DC: National Center for Education Statistics, U.S. Department of Education.
- Madjar, I., McKinley, E. et al. (2010). *Stumbling blocks or stepping stones? Students' experience of transition from low-mid decile schools to university*. Auckland: Starpath Project, The University of Auckland.
- Maharey, S. (2002). *Gateway programme expands to new schools*. Retrieved from <http://www.beehive.govt.nz/release/gateway+programme+expands+new+schools>
- May, T. (2001). *Social research: Issues, methods and process*. Buckingham: Open University Press.
- Middleton, S. (2008). *Tertiary high school report*. Auckland: Manukau Institute of Technology.
- Middleton, S. (2009). *The cost of educational failure*. Unpublished paper, Manukau Institute of Technology, Auckland.
- Middleton, S. (2010). "Bridging a lifetime." Address to the School of Education PostGraduate Teach Education Programme. Auckland: Manakau Institute of Technology.
- Oakes, J., & Saunders, M. (Eds.). (2008). *Beyond tracking: Multiple pathways to college, career, and civic participation*. Cambridge, MA: Harvard Education Press.
- Polesel, J. (2002). Schools for young adults: Senior colleges in Australia. *Australian Journal of Education*, 46(2), 205-221.
- Punch, K. F. (2005). *Introduction to social research: Quantitative and qualitative approaches*. London: Sage.

- Rogers, J., Kahne, J., & Middaugh, E., (2007). Multiple pathways and the future of democracy. In J. Oakes & M. Saunders (Eds.), *Beyond tracking: Multiple pathways to college, career, and civic participation*. Cambridge, MA: Harvard Education Press.
- Rumberger, R. W. (2001). *Why students drop out of school and what can be done*. Paper prepared for the conference, "Dropouts in America: How severe is the problem? What do we know about intervention and prevention?" Civil Rights Project, Cambridge, MA: Harvard University.
- Schwartz, R. B. (2004). Multiple pathways – and how to get there. In R. Kazis, J. Vargas & N. Hoffman (Eds.), *Double the numbers: Increasing postsecondary credentials for underrepresented youth* (pp. 145-154). Cambridge: Harvard Education Press.
- Steinberg, A., & Almeida, C. (2007). Creating pathways for struggling students within a 9-14 system. In N. Hoffman, J. Vargas, A. Venezia & M. Miller (Eds.), *Minding the gap: Why integrating high school with college makes sense and how to do it* (pp. 167-174). Cambridge: Harvard Education Press.
- Stern, D., Raby, M., & Dayton, C. (1992). *Career academies: Partnerships for reconstructing American high schools*. San Francisco: Jossey-Bass.
- Stern, D., & Stearns, R. (2008). Evidence and challenges: Will multiple pathways improve students' outcomes? In J. Oakes & M. Saunders (Eds.), *Beyond tracking: Multiple pathways to college, career, and civic participation*. Cambridge, MA: Harvard Education Press.
- Symonds, W. C., Schwartz, R. B., & Ferguson, R. (2011, February). *Pathways to prosperity: Meeting the challenge of preparing young Americans for the 21st century*. Report issued by the Pathways to Prosperity Project, Harvard Graduate School of Education. Retrieved from http://www/gse.harvard.edu/news_events/features/2011/Pathways_to_Prospersity_Feb2011.pdf
- Taylor-Powell, E., & Henert, E. (2008). *Developing a logic model: Teaching and training guide*. Board of Regents of the University of Wisconsin System. Retrieved from <http://www.uwex.edu/ces/pdande/evaluation/pdf/lmguidecomplete.pdf>
- Tertiary Education Commission. (2003). *Evaluation of the Gateway Pilot*. Wellington, New Zealand: Tertiary Education Commission.
- Trade Training Centres in Schools*. Retrieved from www.tradetrainingcentres.deewr.gov.au
- Tyler, J. Lofstrom, M. (2009). Finishing high school: Alternative pathways and dropout recovery. *The Future of Children*, 19(1), 77-103.
- Vaughan, K., & Kenneally, N., (2003). A constellation of prospects: A review of STAR (the Secondary-Tertiary Alignment Resource). Wellington: Ministry of Education.
- Vocational Education and Training in Schools*. Retrieved from www.deewr.gov.au/schooling/careersandtransitions/vocationalllearning
- W. F. Kellogg Foundation. (2004). *Logic model development guide: Using logic models to bring together planning, evaluation, & action*. (#1209). Battle Creek, MI: W.K. Kellogg Foundation[online]:<http://www2.uta.edu/sswmindel/S6324/Class%20Materials/Program%20Evaluation/Execuitiv.pdf>[April 12, 2010].
- Wholey, J. (1979). *Evaluation: Promise and performance*. Washington, DC: Urban Institute Press.

SECTION 2:
Keynote Speakers

Assessment in a Changing World

Eva L. Baker

*National Center for Research on Evaluation, Standards, and Student Testing (CRESST)
University of California, Los Angeles (UCLA)*

Email: eva@ucla.edu

This paper addresses four interrelated topics. First, the organisation I direct and how its mission connects to the goals and strands of this conference; second, a consideration of the process of assessment, past approaches, current methods, and future designs. This section embodies the substantive theme of the paper. A third set of issues is the challenges we now face in the world of policy and in the face of rapidly emerging technologies. Last, the paper concludes with suggestions for finding the way forward. Were we able to accomplish the myriad of goals envisioned for assessment, it would be truly magic.

Starting with CRESST (the National Center for Research on Evaluation, Standards, and Student Testing), this is an organization that includes a number of complex relationships. First, at UCLA, it is home to the Center for Advanced Technology Systems, which includes the Center for Advanced Technology in Schools, funded by the U.S. government agency, the Institute of Education Sciences. CRESST is also connected to the Center for the Study of Evaluation, which for over 45 years has been a major source of ideas for the evaluation of educational policy and practices, and for scrutiny about testing approaches. CRESST continues that interest in studies conducted for the U.S. Department of Education and the Gates and Hewlett Foundations emphasising the current round of educational reform, a shared vision by autonomous States of goals and assessment, termed the “Common Core of State Standards.” These address literacy, mathematics, and science. CRESST has also been supported by other federal, state, and private agencies looking to improve education through investment in after-school programmes, through improved information systems and reporting strategies, and better support for teacher knowledge and skill development. CRESST has been an important player in educational policy, advising on our National Assessment of Educational Progress (NAEP), Organisation for Economic Co-operation and Development (OECD), Asia-Pacific Economic Cooperation (APEC), and the Office of Science and Technology Policy of the President.

A major component of CRESST work is in the simulation and games areas, and technology has been a central component of CRESST since its inception. Research, development and evaluation have been conducted for entities such as Apple Computer, AT&T, and the Public Broadcasting System (Kids portal), with an emphasis on how technology may be used to improve learning, and as important, how one would provide evidence that it had done so. A major element of research and development has occurred through the auspices of the Office of Naval Research (ONR), a premier research agency in the U.S. government, and through the Defense Advanced Research Projects Agency (DARPA). Our most recent work for these agencies involves the development of strategies to assess the 21st century on the job skills for enlisted personnel and officers, including problem solving and decision making. Another strand of work focuses on the development and assessment of games intended to teach advanced content to young people. In this work, we juggle a number of constructs, including the game developers’ penchants for motivation and fun and our needs to assure that intended learning takes place. We are also engaged in work using technology to address affective goals. For young children, we are addressing empathy, self-efficacy, helping behaviours, and resilience. For adults, we are evaluating systems intended to assist returning veterans with problems of post-traumatic stress disorder and depression. In addition, we have stepped out into the field of medicine, working on simulation tools, curricula, and evaluation strategies intended for medical professionals such as nurses, surgeons,

emergency medicine specialists, and general practitioners. Almost all of this work involves STEM components, especially mathematics and science, and our newest projects, again with support from ONR, will be in the area of intelligent tutoring systems.

So the CRESST portfolio extends across the new social media, classrooms, policy, training, and teaching technology.

CRESST's complexity includes an emerging international mission, involving collaborations with other countries and potential CRESST organizations located in university settings abroad.

Within the U.S., we have many collaborators as well, including researchers from Stanford; University of Southern California; University of California, Santa Barbara; University of California, Davis; Arizona State University; Harvard; Massachusetts Institute of Technology; the University of Colorado; and Carnegie Mellon University, among others. We also collaborate with the private sector, for instance, technology firms such as Intific and Bolt, Beranek and Newman.

Inside CRESST, we have a young and strong interdisciplinary staff, with training in education, psychology, computer science, history, physics, biology, neuroscience, psychometrics, statistics, evaluation, policy, and literacy. We have a strong programme of second-language assessment research, research in the arts, and in the career development area, all generated by interests of our staff.

CRESST leadership must maintain both scholarly and policy interests. They have been awarded many honours, served in prestigious positions, e.g., the Board on Testing and Assessment of the National Academy of Science, and as advisors to governors, education leadership, and to Federal agencies and private charities.

Why this protracted discussion of this organization? Because it will permit the reader to comprehend some of the experiential context from which the perspectives in the paper emerge and allow a stronger evaluation of the value of commentary. It also permits the connection to strands in this effort that are relevant to the agenda of the meeting and to CRESST's long-term interests. These include standards-based assessment design and validation, college, further education and workplace-ready knowledge, skills, and attitudes, underperforming students, measurement advances, explicitly design methodology for technology-based tests, games and simulations, technical quality for extended performance tasks, professional development and value-added teacher evaluation, programme and policy evaluation, and the generation of tools to make all these goals a practical reality.

Assessment: What Is It For?

When we think of assessment it is often about how much knowledge in a content area, such as mathematics, students have obtained, often measured by some standardised test. In the present, we are confronted with an expectation by many students for more personalised learning, addressing their particular interests, needs, and skills, whereas many societies are similarly interested in traditional learning, as well as the importance of continuing shared, strong, and resilient values. These ideas are in some conflict and affect how new assessment will develop.

Depending upon how far back we go, we can remember a time when assessment was not a major factor in public policy. In classrooms, most tests were either developed by teachers or cadged from the back of each chapter in the textbook. Sometimes the tests were strongly connected to classroom activities, but often the purpose of the test was to have some "objective" indicators of student performance to give marks at the end of the term. It was also seen as a method of motivating students, and in times gone by, probably worked. When people talked about tests they typically used quick metaphors – multiple choice was thought to be most objective and relegated to memorised facts and "subjective"

tests were those that required performance, such as an essay, paper, or project to be judged (privately) by teachers. Now, of course, the range and complexity of test formats have multiplied because of the availability of technology for administering, monitoring, and even scoring complex or extended performance.

In fact, the purposes of tests have greatly expanded beyond the motivation and grading perspective. Tests became used for admission and placement decisions in higher education, then private schools. Tests were also used to certify adult accomplishments, such as medical board examinations and for other licensure purposes. Various methods have been used to classify purposes of tests. A simple one is to determine whether the test results imply sanctions, rewards, or punishments, for individuals or for institutions. Examples of tests for this purpose would be those used for personnel decisions, diploma granting, and school accountability. This use is compared to tests, which are to help the respondent, that are for general rather than specific purposes, or may allow students to gain feedback and directions for improvement.

Two major changes provide tension in testing and assessment prospects. Among the changes that have impinged on the world of testing is its use as an instrument of public policy. Instead of having tests measure the impact of instruction, recent decades have seen the use of tests to delimit the curriculum itself. The effects of test-driven teaching, in many places in the United States, are less interesting, challenging skills and content are being taught, and, in some cases, drilled.

Second, an opposing force has developed. Through the efforts of cognitive science, it has become clear that serious learning requires extended engagement by the student, if it is expected that skills and knowledge will be retained and applied in the future.

The learning perspective is clearly to be preferred. If learning is taken seriously, and assessment represents a fair and adequate measure of desired learning, then assessments will have to be redesigned. In the face of a knowledge explosion and the heretofore comment about students' desires for interest-driven outcomes, it will be important to rethink the practicalities of creating and recreating assessments in the future.

If we focus on learning-driven assessments, they will require design of a series of components. Table 1 lists those components.

Table 1: Learning-Driving Assessment Design Components

-
1. Cognitive Skills (21st century, of course)
 2. Content
 3. Criteria to judge quality of performance
 4. A specified range of situations
 5. Internalization and transfer across situations, constraints, and combinations of content and cognitive skills
-

Cognitive Skills

In considering assessment redesign, two principal issues are economy and validity. The process, we believed, would better start with developing good definitions and examples of families' cognitive skills. Then this could be embedded in content at different types or levels. The problem of describing and measuring the skill level would be more difficult than finding the content to be used. Thus, late in the 1990s, CRESST developed a short set of cognitive skills to model assessment design. They are included in Figure 1.

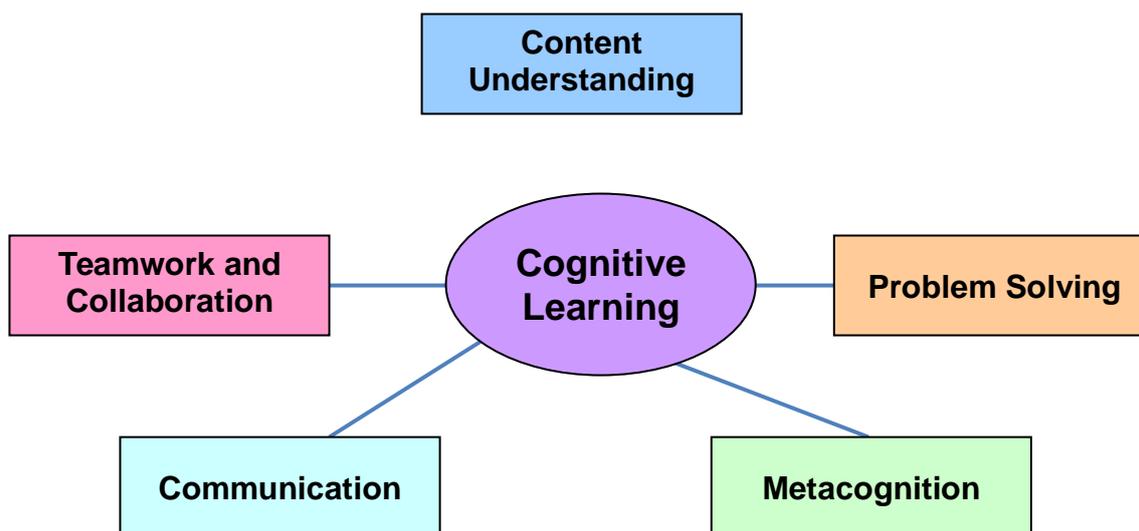


Figure 1: Cognitive Skills for Assessment Design

In this model the content, whether declarative, procedural or strategic, would be the stuff into which the cognitive skills were placed: problem solving, communication, metacognition, and teamwork and collaboration. This approach was used by CRESST to develop statewide trials of history examinations in Hawaii and assessments of literacy for the Los Angeles unified school district. In both cases, costs were an order of magnitude less than competitive prices.

Ten years later, a new set of cognitive skills was constructed from the old. The major differences were the existence of new data (Endsley, 1995; Klein, 2002; Mayer, 2010; Sweller, 2003) about the utility and generality of the skills. A second important variation is that the skills are thought to be 21st century because they represent very seriously the importance of learning to transfer and generalise over unpredictable situations. This requirement is obvious as the pace of change accelerates, both in knowledge and in potential career options. The ability to adapt, to imagine success in a variety of arenas, and to show flexible intellectual skills in a variety of particular content areas is key. In Table 2, 21st century cognitive skills are listed.

Table 2: 21st Century Cognitive Skills

-
- Adaptive, complex problem solving
 - Situation awareness and risk assessment
 - Decision making
 - Self-regulation
 - Teamwork
 - Learning to learn
 - Communication
 - Conceptual, procedural, and systemic learning of content
 - Application and transfer
-

Very brief examples of outcome measures are provided in Table 3.

Table 3: Examples of Outcome Measures

-
- Design of outcome measures
 - Content understanding
 - Concepts, principles, descriptions and predictions of scientific phenomena (e.g., identify the forces imparted on a ball)
 - Problem solving
 - Design a solution under constraints (e.g., program a robot to follow a predefined path)
 - Transfer
 - New situations different to those used in learning, requiring similar problem solving processes, content expertise, or SEL (e.g., using analysis and reaction to bullying that applies to another anti-social behaviour)
-

Although there is rich literature related to each of the proposed skills, let me move along to the problem of content.

Content Models

For many years, content was conceived as a broad construct: math ability, American history knowledge, and biology. Under that construction, it would be fair to select a wide range of examples or specific content because most of any selection would cohere with the construct of ability.

We have taken another approach (Baker, 2007; Baker, Chung, & Delacruz, in press; Chung, Delacruz, Dionne, & Bewley, 2003; Iseli, 2011; Iseli, Koenig, Lee, & Wainess, 2010, Koenig, Lee, Iseli, & Wainess, 2009) where the content domain is explicitly bounded and described, and subsumed content is represented. We have chosen to use ontology to represent knowledge in a network environment. Where does the knowledge come from? From groups of subject matter experts who give their best interpretation of standards expected for student learning. They also hypothesise alternative routes through which the content, linked to cognition, may be best learned. The purposes of the content ontology are many: (1) to provide a crosswalk or method to get from general, verbal descriptions in standards to explicit learning goals; (2) to clarify relationships among knowledge components; (3) to guide design and review of instruction and assessment; and (4) to provide transparent pictures for teachers and students.

In Figures 2 and 3 are representations of how ontology functions.

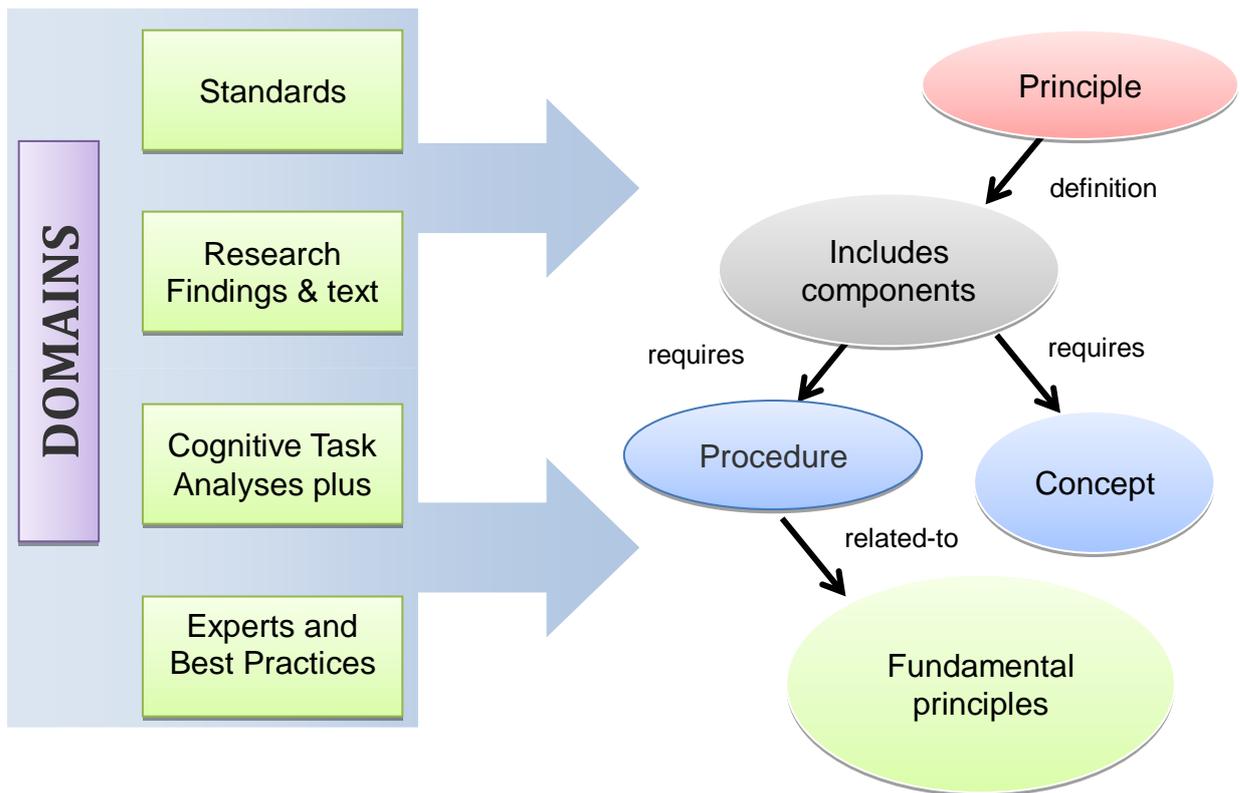


Figure 2: Ontology Design

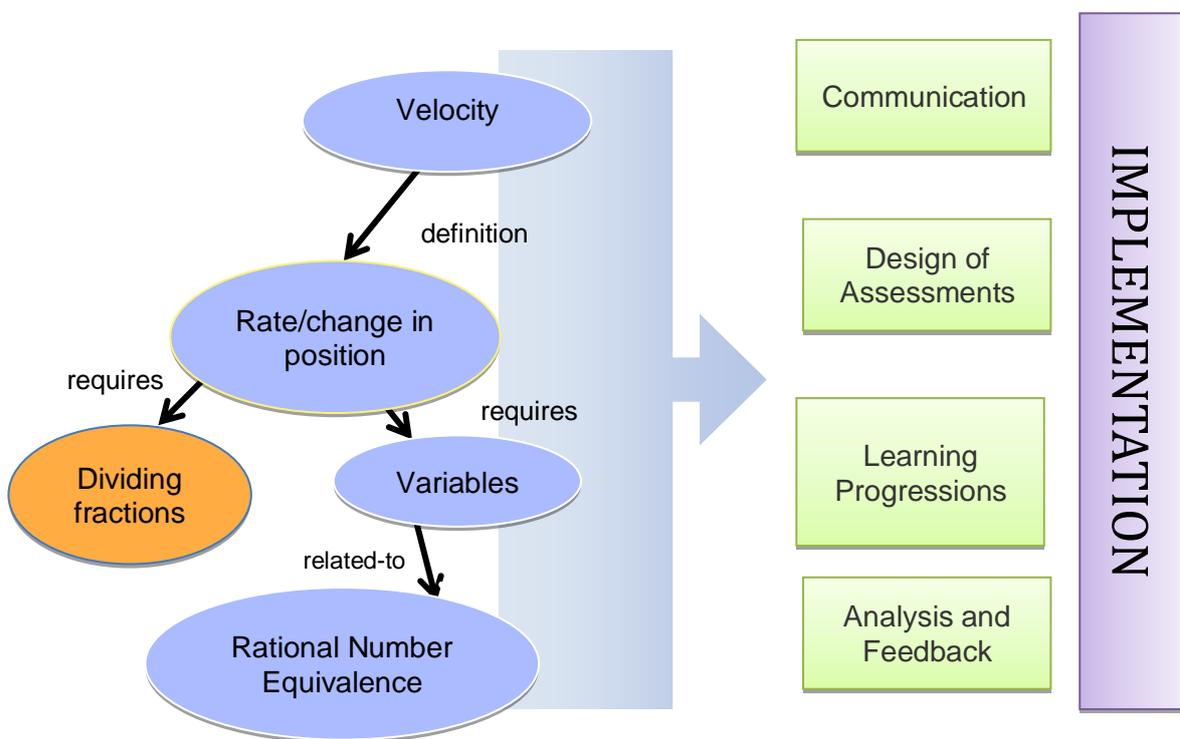


Figure 3: Ontology Use

Figure 4 gives a simplified version of an Algebra ontology to show the key relationships. Important ideas are identified by how many connections in the graph are made to any node. Also relationships between pairs of nodes are made explicit.

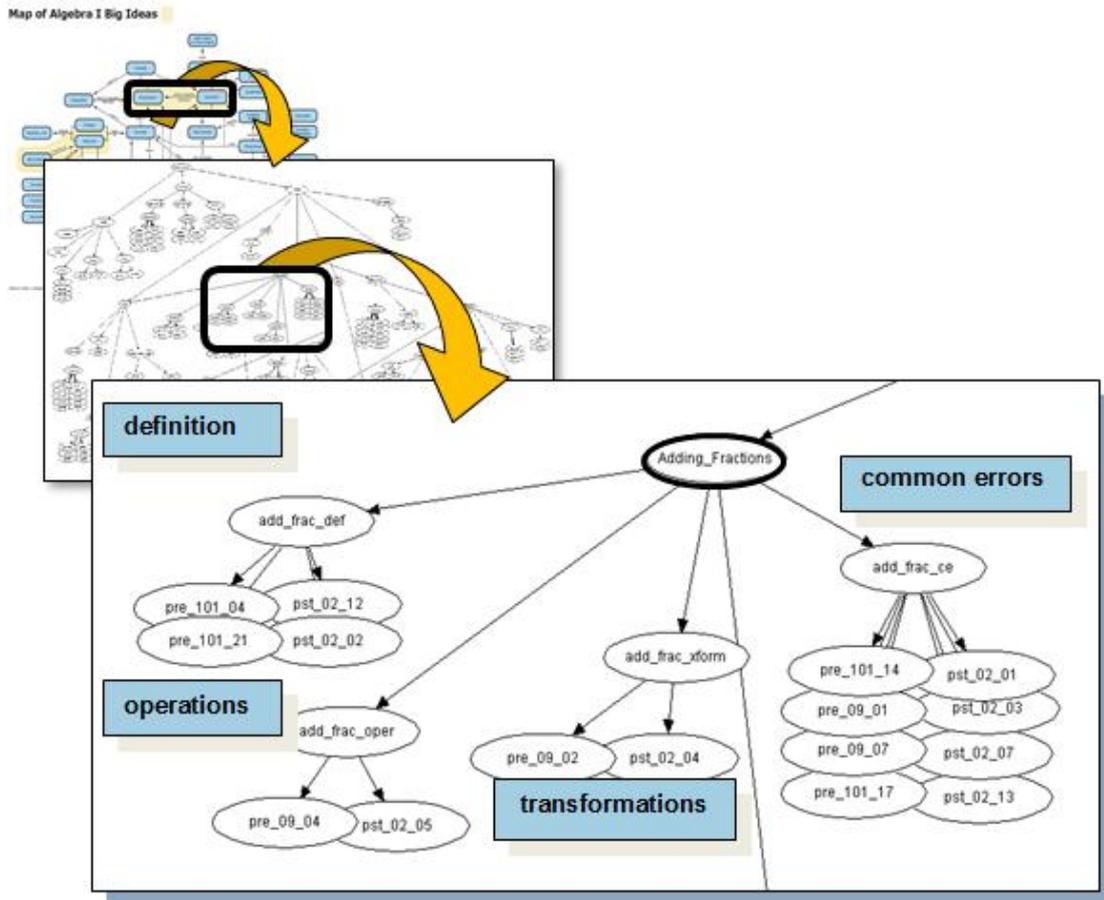


Figure 4: Ontology Map of Algebra Content

Of particular interest is the way the map can be used in data acquisition. The ontology can serve as a framework in which to place student responses to relevant test items. These responses will depict the actual relationships found in student learning rather than the hypothesised sequences. Based on success on key tasks, a statistical analysis can re-order the best sequence for either the entire group of students (optimising) or could identify preferred sequences for students who start with particular patterns of strengths and weaknesses. Thus, in a kind of machine learning, the computational changes in student sequences will be exhibited and may be used to revise or rethink design plans for assessments (particularly formative assessments used during instruction) or the instructional sequence itself.

Model-Based Assessment

Thus the development of learning-driven assessment will start with a model of learning based on research on cognitive skills and best knowledge of experts related to content. The cognitive components will be reusable, because common templates, or frames for setting up problems can be shared across grade levels and some content areas. At CRESST we have also applied reusable scoring criteria to explanation across subject matters. For instance, the rubric will include the use of major principles or themes, the use of concrete examples taken from source materials or documented in other ways, references to prior knowledge that links to claims and argument, as well as standard style and language requirements. To use these rubrics, it is clear that they must be embedded in content. The principles must deal with the task content under assessment, whether it is economics, history, literature, biology, physics, or mathematics. Similarly, the knowledge, concrete examples or illustrations and prior knowledge, are the facts, concepts, or

procedures relevant to the particular subject matter (or interdisciplinary task) under review. These kinds of rubrics then save enormous time in development and validation. They only require training by teachers or other markers to use them, and retraining for a different topic or subject matter is relatively quick.

Technical Quality

For learning-driven assessments to work, they must have evidence of their technical quality. When we are looking at tasks that are multi-stepped, extended in time, or involve a series of cognitive skills, we are out of the standard psychometric ways of determining the quality of tasks. For instance, reliability approaches must be modified to take into account the dependencies that occur while completing a long task, with the understanding that some tasks take so long it is impossible to test one student on many of them. For validity, the challenges remain. One is that a single assessment may have multiple purposes, for instance, to rank students, guide teaching practice, provide feedback and help in improving for learning, and give data for accountability systems. It is obvious that all such uses need validity evidence that allows one to infer that the assessment is providing accurate and useful knowledge. However, in practice such evidence is rarely available.

More Validity Concerns

As standardised tests move to systems of Qualifications or badges requiring the accomplishment of a series of tasks, some difficulties not unlike those encountered when portfolios were developed will occur (Gearhart, Herman, Baker, & Whittaker, 1993; Herman & Gearheart, 1998; Herman, Gearhart, & Baker, 1993), both involving accuracy of judgment and fairness. As a first cut, some expert-novice comparisons may be investigated, where people thought to be proficient are compared to those who are known to be novices or individuals just at an early point of acquiring competence. Quantitative and qualitative differences in performance can point to specific remedies for the design and marking of the qualification.

While there are numerous guidelines for validity (Messick, 1989), the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education, 1999, its revision soon to be released), they focus on the argument made that the purposes of the assessment have been met through the process used. Minimum requirements require the expert judgment of tasks and their relevance to an ontology or other content depiction. Second, during the course of development, think-aloud protocols should be used to ensure that the respondents are using the types of thinking skills intended to be measured. Ideally, large-scale trials occur prior to application of the test or assessment in practice. This desire is less frequently met as schedules for testing have accelerated while costs allowed for development have dropped.

When components of validity are addressed, for measures intended to assess learning, instructional effectiveness, or school efficacy, then it is clear that there must be evidence that the measure changes because of instruction. The alternative is measures of innate talent, ability, or out-of-school influences. While levels of economic and educational backgrounds will continue to matter, it is most important that judgments made about learning in an educational context are empirically tied to the changes occurring in learners.

Technology Challenges

Finally, more challenges will occur in technical quality as a function of the use of multiple technology platforms for administering and scoring performance. At this point, there is great difficulty in ensuring who is actually taking the test, if administered in a non-

proctored setting. Log-on procedures may be inadequate, because the actual performer of the tasks could change. Approaches using cameras, and advanced analysis of linguistic and writing preferences may one day be able to verify the test taker's identity. There are numerous, but still unsatisfying approaches to the online measurement or automatic scoring of open-ended, complex student performance. These involve questions of whether the scoring approaches can be gamed, e.g., using many long sentences, or whether they simply measure surface or grammatical features, and suffer from evidence that they address the meaning of the student work. Moreover, such systems are open to the increasingly prevalent hacking of systems, systems including financial and defence information that have far greater protection than educational systems. Naturally, solutions will be found to these difficulties, but in the short run, they cause some worry.

Policy and Future Development Challenges

To do a good job of education, our measures must be lined up with our goals, not the other way around. If we are interested in cognition and content, we also should be concerned with the reality of the 21st century, ability to adapt and transfer, ability to self-manage, and ability to work with others. We are also interested in students' affective states, beyond the simple surveys of whether they "like" science and would be interested in pursuing it as a study or career. We are interested in risk taking and entrepreneurship. We are concerned with the right kinds of character, morality, and emotional development. Measuring these in a non-clinical setting will be very difficult, partly because of privacy concerns, and partly because no-one yet has achieved a scalable approach to measuring and accepting a wide range of values across individual students. But if school is to be more than the acquisition of small bits of knowledge or inculcated procedures, we as professionals will need to move ahead.

The Way Forward

Clearly, as a researcher, I will support the need for more R&D in assessment. It is critical as our goals change, our discretionary resources recede (and then expand) cyclically, and the world changes around us, we must find faster ways to keep apace.

We will inevitably confront the issue of unconventional, or personal educational goals, which places us in the tension among free choice, some flexibility, and institutional control. We will need to use better designs, data capture, analyses, and reporting techniques. The current surge of interest in populist approaches, e.g., crowdsourcing or popular ratings (stars and likes) may be fine for choosing restaurants or judging a picture on a social media site, but are surely inadequate when the goal is verifying student learning. All of our efforts will be undertaken in a changing environment, be unstable, and requiring iterations. We can succeed if we base our plans on clearly developed, usable knowledge, and verify the validity of our approaches with both tried and true and new (brain imaging) techniques.

Success will inherently depend upon our own ability to cooperate, share, collaborate, and develop a range of approaches suitable to our students, our settings, and our value structures.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Baker, E. L. (2007). Model-based assessments to support learning and accountability: The evolution of CRESST's research on multiple-purpose measures. *Educational Assessment* (Special Issue), 12(3&4), 179-194.
- Baker, E. L., Chung, G. K. W. K., & Delacruz, G. C. (in press). The best and future uses of assessment in games. In M. Mayrath, D. Robinson, & J. Clarke-Midura (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research* (pp. 227-246). Charlotte, NC: Information Age.
- Chung, G. K. W. K., Delacruz, G. C., Dionne, G. B., & Bewley, W. L. (2003). Linking assessment and instruction using ontologies. *Proceedings of the I/ITSEC*, 25, 1811–1822.
- Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors*, 37, 65-84.
- Gearhart, M., Herman, J. L., Baker, E. L., & Whittaker, A. K. (1993). *Whose work is it? A question for the validity of large-scale portfolio assessment* (CSE Tech. Rep. No. 363). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing.
- Herman, J. L., & Gearhart, M. (1998). Portfolio assessment: Whose work is it? Issues in the use of classroom assignments for accountability. *Educational Assessment*, 5(1), 41-55.
- Herman, J. L., Gearhart, M., & Baker, E. L. (1993). Assessing writing portfolios: Issues in the validity and meaning of scores. *Educational Assessment*, 1(3), 201-224.
- Iseli, M. (2011). *Ontology development: Overview and example* (Draft CRESST Whitepaper). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Iseli, M. R., Koenig, A. D., Lee, J. J., & Wainess, R. (2010). *Automatic assessment of complex task performance in games and simulations*. (CRESST Report 775). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Klein, S. B. (2002). *Learning: Principles and applications* (4th ed.). New York: McGraw-Hill.
- Koenig, A. D., Lee, J., Iseli, M., & Wainess, R. (2009). *A conceptual framework for assessing performance in games and simulations*. Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC).
- Mayer, S. (2010). *Problem solving ontology* (CRESST Whitepaper). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (pp. 13–103). New York: McMillan.
- Sweller, J. (2003). Evolution of human cognitive architecture. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 43, pp. 215–266). San Diego, CA: Academic Press.

National Education Standards for New Zealand: A Research Agenda

John Hattie
University of Melbourne

Email: jhattie@unimelb.edu.au

Abstract

The degree of implementation and effects of the National Standard policy provide an excellent opportunity to evaluate a model based on teacher overall judgements rather than more typical test-based accountability models. This chapter outlines seven research programmes, some commenced, and their progress to date.

New Zealand has developed a laboratory for excellent research. We punch above our weight in many domains of research and my argument in this paper is that National Standards offer an opportunity to lead the world in how an accountability system based on teacher professional standards can positively impact teaching and learning – if indeed it does so positively. The alternative, test-based accountability systems, is worse with many negative wash-back effects. To develop and execute such a research plan would mean we need to move past the advocacy and anecdotal, and join as a community to systematically pool our research expertise to address the many opportunities of this policy.

New Zealand is a country that has, on average, high performance in the international achievement league tables (as per PISA, PiRLS, and TiMSS), great dispersion in the spread of achievement, and regarded by many as a very impressive educational system. The policy on National Standards can thus be seen as potentially a 'value added' policy (or not) and evidence of its impact could be subject to some of the most exciting research in the world.

I recall the excellent laboratory that New Zealand provided for the evaluation of vouchers and decentralisation that was so ably written about by Fiske and Ladd (2000). The two major foci of this conference, National Standards and NCEA, provide another natural laboratory – both assessment related, and both potentially changing teaching and learning and placing New Zealand again at the forefront of educational research.

In this session I wish to outline seven research programmes, and not only review what has occurred so far, but recommend that academics and educators sustain and nurture these research programmes and make sure that the National Standards policies and implementation are thoroughly the subject of research. If not because the policy is controversial; if not because there are so many opinions and blogs rampant about criticism usually based on prior world views, anecdotes, and advocacy; if not because as education researchers we can have an influence on the policy direction and implementation; but because we are optimally placed to conduct research, show its power and value, and to bring many types of research design to a national issue.

In many ways like National Standards, the NCEA seems to have been a great innovation handicapped by a poor initial implementation. I would suggest that if the high school teachers of New Zealand had been told before the introduction of NCEA that they would undertake the greatest transformation as to how they teach, this would have been greeted with laughter, disdain, and strikes! I would argue that there is much evidence to show that there have been dramatic changes in the nature of teaching, certainly higher levels of student retention, and now one of the more important concerns in New Zealand is a function of the success of this teaching via NCEA – how to cope with the number of students completing regular high school not wanting to go to university or enter a trade; we have no community colleges, liberal arts colleges, no place to continue learning for the sake of learning. Now there is (another) great research topic.

Research Issue #1. The Nature of the Policy – Policy Research

Probably all of us in this room know about Tomorrow's Schools, but note that "Tomorrow" has been the past 21 years. We have sucked out all the benefits of this policy, and its many problems are transparent but the fear of returning to the bad old days of Aunty Ministry and Daddy Regional Directors seems to stop the discussions of moving forward (Hattie, 2009a, 2009c).

The National Standards policies test the core notions of Tomorrow's Schools, and much of the resistance is to this 'test of the core'. In a simple but not incorrect summation – Tomorrow's Schools allowed the government and Ministry to decide policy, but gave autonomy to the schools to implement it. The aim was to increase the professionalism of teachers and the local respect and involvement in schools. National Standards does not give such autonomy to the schools, hence much of the reaction and critique. For many, a dagger is being thrust at the heart of professionalism of schools.

Thus, the challenge for policy researchers – How can a policy (such as National Standards) that seems apposite the current zeitgeist (TS) exist *without* addressing the zeitgeist. I have recommended elsewhere that there be a national inquiry or ministerial review of schooling leading to a new metaphor or basis of educational policy – post-Tomorrow's Schools – in 1944 the Thomas Report, in 1964 the Currie Commission, and in 1989 Tomorrow's Schools. So every generation we have relooked at the fundamentals of education. We did not do this when we implemented National Standards and this has led to contradictions about current policy.

The notion underlying TS is "let 2800 flowers bloom" or to mix the metaphor, let's let the cream rise to the top. We have seen the dilution of the more extreme parts of the competitive model, although it is still present. National Standards *could* reduce one of my major criticisms of the residue of TS – it could increase cooperation between schools but it is hard to see this in reality as each school interprets the policy, creates its school reports, deals with its mandates, and worse – differentially interprets the meaning "progression" inherent in the National Standards.

When I review educational policy in other developed countries there is an emerging trend that should worry New Zealand and is well worth making a sustained research topic – the increasing number of voters who have no child involved in schooling (or involved in state schools); hence the long tradition in New Zealand of supporting state schools may slip. I note in Western Australia, the number of students going to state schools has slipped to about 50%, and their new policy of creating (now about) 150 public-private schools within the state system further erodes the voting bloc! Are we witnessing, in National Standards or the reaction to National Standards, the first major move in New Zealand towards a differential public-private state system, an erosion of the national state system of schooling?

Indeed, the policy researchers in New Zealand should be abuzz with debate. Note the effect of one letter from some academics on policy debate in National Standards about two years ago! (Thrupp, Hattie, Cooks, & Flockton, 2009). A major consequence of this conference would be to prepare a plan for conducting such research, a special issue for a Journal, and an exciting influence that academics can have on the national debate – and be so informative for international audiences.

Research Issue #2. National Standards: Fixing the Tail

One of the claims for the existence of National Standards is the '20% tail' and here is a great research topic. Where did this 20% come from? In one claim, given that so often we discover a normal curve underlying achievement and ability (not always, but this is an empirical finding) then there is always a tail! When "standards" are set they are often "aspirational" or some like term; hence it will mean that it is very difficult to get "all" students over whatever hurdle is set. I remind you of the Blair government aspiration of

getting 80% of students to their curriculum standards – they saw 20% as not getting there. (That they moved from the mid 50s to late 70% was interpreted by many that they had failed!)

We know from the many speeches throughout New Zealand that we are a “high quality, low equity” country, but then most countries I have been in say the same. As I have argued elsewhere, this leads to poor metaphors, poor analyses, and poor explanations (Hattie, 2009b). The use of terms like “the tail”, the underachieving, the bottom 20% has led to many millions of dollars being spent on solving the wrong problem – leading to no change in the bottom 20%.

There will *always* be a bottom 20%; there will always be a tail; and more importantly, you cannot use achievement tests alone to detect an under-achievement student. With Fletcher, I published a book earlier this year on the use of intelligence tests – one of the arguments is to revisit these methods as a way to see discrepancies between what a child is able to do relative to what they do, but know this is highly unlikely (Fletcher & Hattie, 2011). Another way is to set standards for students and then see the gaps between where they are and where they should be. There are two faults of this thinking, however.

First, the distance between current achievement and potential achievement is different for each student. When reviewing the distribution of reading achievement – based on 1m students in the asTTle data base, there is a normal distribution (no surprise). This is the case for the distribution of Pākehā, Māori and Pasifika. The problem is those students – across all parts of the distribution – can be below where they “could be”. There are, for example, as many Māori students behind Pākehā students above as well as below the average. Where are the programmes, other than Bishop’s Te Kotahitanga to address the concerns? A fetish with the tail misses this problem.

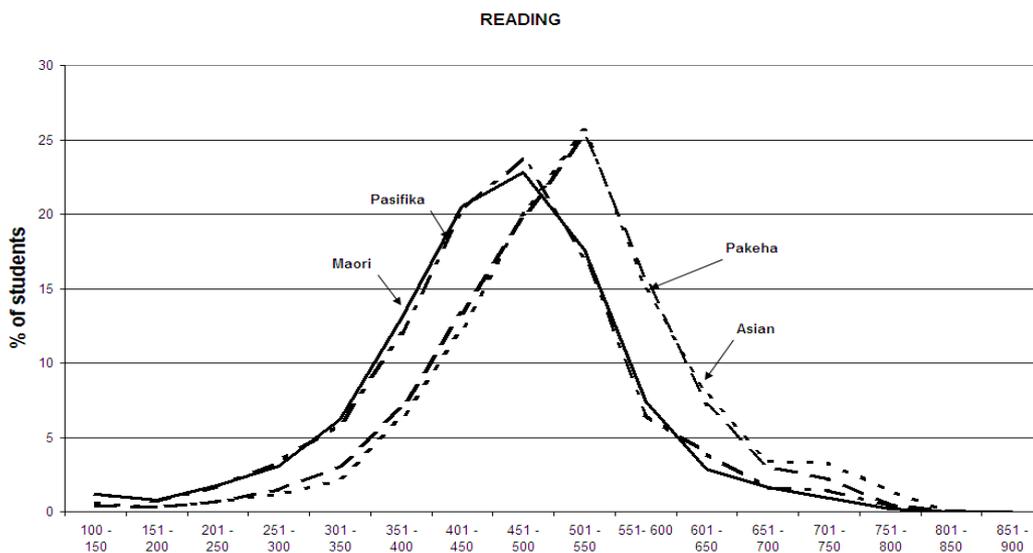


Figure 1: The Distribution of Reading Achievement for Pākehā, Māori, Pasifika, and Asian

Second, the National Standards are based on “years” – and in my view this was the worst decision made in the details of this policy. Using “years” presumes all students of the same age year can perform, can gain, and can move towards the same expectation. If we know anything about schooling it is that “years” (e.g., Year 4, Year 5) means that there is a spread of four or five years in any one age group! It also flies directly in the face of the development of our current curriculum which is NOT based on years but on a deeper notion of development. While there is a “year” base in National Standards and a progression base in curriculum there is an absurdity that schools are asked to reconcile – and the best way is to ignore one of these mandates. This needs much more debate about how to reconcile these competing bases.

Research Issue #3. The Problems that National Standards Aim to Address; Goal-Free Evaluation

Scriven (1972) introduced the notion of goal-free evaluation, and such a notion is well worth pursuing in the research on the problems that National Standards are meant to resolve. While I acknowledge that many political manifestos have at their core a plea for voter acceptance, National Standards does provide a fascinating case study for goal-free methods.

Yes, there were reasons promulgated by the National Party when introducing National Standards, but Scriven’s argument is that these should be ignored and instead we need to ask what has been affected – and then these are matched with the goals of the promulgators. Of course, the avowed reasons for introducing National Standards may well be correct (or not), for example, the claim in their 2008 Manifesto that “according to ERO, most primary schools are generally ineffective at using school-wide information to improve achievement.”

(a) Retention

We know from the outset the claim was that the school system needed to prepare students for passing Level II NCEA – a laudable aim, and close to the US claim about “college readiness” which has so driven their policies. From the NCEA annual report, we know 27% do not attain Literacy or Numeracy units in Level 1; about 60% attain Level II (although almost halved for Māori and Pasifika = 37%). [And note the success of Selwyn College moving from about 30-40% pass in numeracy to 90%+ by setting higher expectations/standards for their teachers and students. Standards can drive performance upwards!] This did not “grab” as a policy imperative but surely this is a fundamental concern. If, as Levin, Belfield, Muennig, & Rouse (2006) have shown, it is not achievement that is the best predictor of health, wealth, and happiness in adult life, but retention in school, and as many students make the decisions about staying in schooling in the middle years, we have a major issue for the future of New Zealand.

I would argue from the evidence provided at this conference that the NCEA has led to major successes in increasing retention rates. Over the past seven years there has been a 10% increase (from 70% to 80%) for Pākehā and Asian, 15% increase for Māori and Pasifika (from 45% to 60% for Māori and 38% to 53% for Pasifika) (See Figure 1).

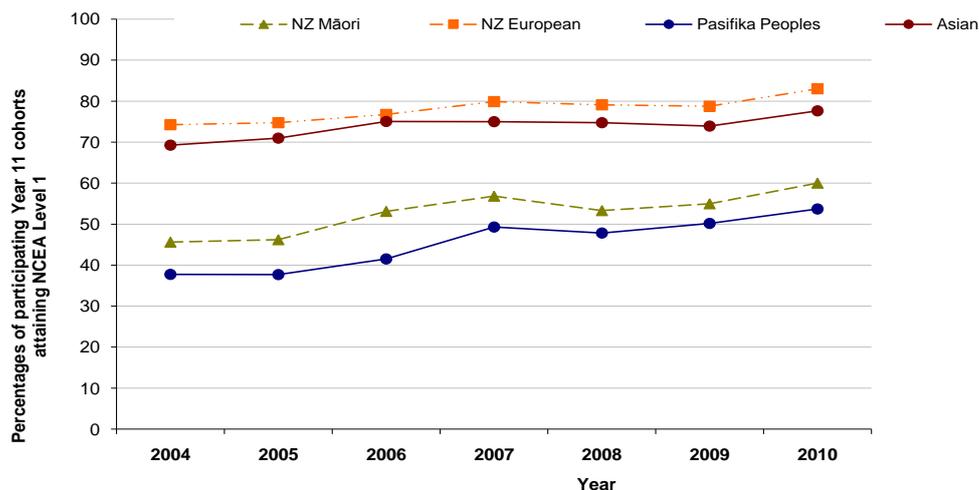


Figure 2: Percentages of participating Year 11 Asian, New Zealand European, New Zealand Māori and Pasifika candidates attaining NCEA Level 1 from 2004 to 2010. All standard errors are less than one percentage point

What evidence would we seek to see any consequences of National Standards on retention rates in school? If only we had asked this question at the outset of the implementation of NCEA. We need a research plan, a research group empowered with providing this evidence – using a goal-free design. Maybe with the amalgamation of ERO and NZQA there could well be an excellent research group asked to report publicly on evidence of impact of National Standards and NCEA.

(b) Growth and Expectations

One major research request is to turn one of our major issues on its side. So often we seek evidence of impact on proficiency or attaining the National Standards – how many more students are attaining National Standards – and this is indeed very worthwhile. As noted above, however, we will never succeed with all students if these are the criteria. If we ask, first, what proportion of students gain at least a year's growth for a year's input then we can have a higher success – and indeed every child (and their parents) should be entitled to at least a year's growth for a year's input.

My Visible Learning work (Hattie, 2009d) shows that an average effect-size growth per year is about $d > .40$; and this is the finding from analysing the aTTle data base, and now I am in Australia also the average based on the NAPLAN data. The beauty of this measure is that this is reasonably independent of socio-economic status of the students – all can achieve a year's growth (regardless of starting point) and a goal well worth aiming for in the New Zealand National Standards debate. Every school can be asked to show the percentage of students gaining $d > .40$ per year, and use this as the benchmark for evaluating the effects of National Standards.

Otherwise, the percentage attaining National Standards may be more correlated to the housing prices; there may be (as in Australia) a drop in the performance of those above the average (McGaw, 2011), and other well-known NCLB effects (see Baker & Linn, 2004; Darling-Hammond, 2009; Linn & Baker, 2002).

(c) The Presence of OTJs

The emphasis on OTJs (overall teacher judgements) is one of the unique aspects of National Standards in the international community. Elsewhere, tests dominate the discussion, whereas here in New Zealand the aim is for tests of multiple forms (and formats) to be inputs into the OTJ along with many other professional judgements.

By chance, we conducted a research study on the expectations of primary teachers on estimating student achievement in New Zealand. We asked hundreds of teachers to be involved in estimating the “equivalent year” for thousands of items we were importing into the aTTle item data base. We then collected the actual evidence of the students on these items – on average primary teachers underestimated what primary students can do by about one level! Is this a (positive) effect from introducing National Standards, and I note Peter Tait's (2011) work at this conference on this topic.

A worthwhile research plan would be to evaluate the effects of National Standards and OTJs in raising the expectations of teachers and then watching the subsequent student achievement rise to these expectations. Christine Rubie-Davies (2006, 2007, 2010) from UoA has a solid research programme on expectations and how teachers tend to have high or low expectations for all students and well worth making a national study of National Standards and expectations.

(d) A Common Conception of Progress

We presented a series of papers on the state of New Zealand education (Leeson, Brown, et al., 2005, Leeson, Hattie, et al., 2005, Reddish et al., 2006). These were sent to every school in New Zealand and outlined what we saw as the fundamental issue for our schools – teachers need to have a common conception of progress. For example, in Writing there was a flat line in the average performance from age 8 to 16 and this finding

has led to some excellent research and changes, led by Judy Parr among others (Parr, 2010). I showed how, for the last three years of primary school, there was a flat line in reading achievement, to then move upwards when these students went through high school – most likely due to assuming and thus ceasing to actually teach reading strategies and subject matter vocabulary in later primary school (Hattie, 2009).

We conducted standard setting workshops in Reading and Mathematics and found that teachers have a very good and common understanding of levels, but the variance can be enormous. It took a 2-3 hour exercise to bring this variance to a very acceptable level (using a modified Bookmark method), but without such moderation the interpretations about National Standards are likely to be too disparate. I am sure Val Klenowski (2011; Connolly, Klenowski & Wyatt-Smith, 2011) will speak to this issue, as have others who have been concerned with moderating teacher judgements (see Black, Harrison, Hodgen, Marshall, & Serret, 2010). Like in the NCEA, moderation is at the core of the success of National Standards, but the current methods of social moderation (which have failed everywhere else) are insufficient, and more defensible, and research-based methods need to be considered.

Creating a list of problems that National Standards can affect, and then commissioning research on these and the effects and consequences of National Standards, seem minimal. For example, such a list could include:

1. How to enhance teachers' expectations about students' progress and potential
2. Progress – How to assess progress (in a defensible manner across teachers and across schools)
3. Common conception of progress
4. Closing the curves
5. Student assessment capacities
6. Teachers' summative judgements (OTJs)
7. Using targets and enhancing expectations of change
8. Retention in secondary and creating schools as inviting places to learn
9. Schools' self-evaluations of success and progress
10. Schools' transparency of outcomes
11. Engaging in challenge – leading to higher pass rates at Level I into II
12. Increasing % passing literacy and numeracy at level I
13. Developing the best PD for teachers and school leaders they have ever experienced
14. Making the PD providers work closer with schools to create annual growth of $d > .40$
15. Winning the hearts and confidence of the implementers
16. How to gain more involvement from teachers in resolving the problems; How to win their hearts and minds.

Research Issue #4. The Implementation – Monitoring Evaluation Research

Barber (Barber et al., 2010) has outlined a powerful model of change in Education – until the unfortunate title of Deliverology. Fullan (2011) has produced many treatises on this topic, and one of the most common findings is that any change must win the hearts and minds of those implementing the innovation.

Take NCEA, a major motive for many teachers during the inadequate early implementation was to “not harm our students” and many teachers spent an inordinate number of hours redesigning their teaching and assessments – indeed the criticism could well be that they did spend such time – but not with each other. This as we now all know led to major moderation problems.

Compare the list of implementation concerns of the NCEA:

- It was rushed but at the almost last minute it was delayed a year
- It was very much about ideology relating to curriculum
- There were many early problems in the NCEA
- There are unacceptably low levels of assessment reliability
- Limited moderation and lack of consequences for schools making incorrect internal assessments
- The lack of comparability of student assessment between schools and subjects and of qualifications over time
- The use of only four grades
- Difficulties in establishing the authenticity of a student's work
- Teacher workload and assessment overload
- Many looked for alternatives with (often unsupported) claims about their superiority (e.g., Cambridge, IB).

These issues (as well as a scandal) led to a Ministerial group for Scholarship then NCEA chaired by an independent person from outside government (Hawke), and while I have been on this group from the outset, I would regard it as having recommended many initiatives and recommended pausing on others. The major effect has been, and remains, seeing that the underlying issues for Scholarship and NCEA were fundamental measurement not curriculum issues. When will we realise this in the National Standards debates?

Now compare this list to the current issues with the implementation of National Standards. Only one difference – it has not been delayed a year!

- The evidence from ERO (2010, 2011), and the MoE independent research (Thomas, 2011) points to the slower uptake, and there is an uptake. Many schools are taking initiatives to include National Standards in their planning and teaching, many resisting for many reasons, and some just keeping going with minor modifications; again, a perfect natural laboratory for research.
- I have seen the Ministry's program logic – but where are the program logics in schools and how are these related to enhanced outcomes for students – how are schools relating their degree of implementation of National Standards to changes in teachers/teaching and thence learning (or not)? There is a great series of studies to ask about evaluation monitoring.
- From Cuban's (1993) work, there is evidence that the nature of teaching has not changed much over the past 200 years. We are resistant to new methods, not because they are better or worse, but the current model has worked well for most. My impression before I left New Zealand, and from what I saw while on the Independent Advisory Group was, at best, a very thin whitewash of paint being applied in the name of National Standards so that the grammar of schooling does not change. Is it?
- The greatest first change has been to school reports. Peddie and I noted 10 years ago (Hattie & Peddie, 2002) that the majority of New Zealand schools sent home reports showing that 95%+ of students were achieving well, putting in effort, and a pleasure to teach. Koefed (2009, Koefed et al., 2011) showed this had barely changed, and I am delighted she is directing a team to research this round and make recommendations, particularly in light of her research on this topic (Koefed, 2010, 2011).

In terms of research on program logic and National Standards, however, I will not hold my breath, until there is evidence of changes in teaching – otherwise there are unlikely to be changes in learning.

Research Issue #5. The Effects in the Schools – Success Based Research

Brinkenhoff (2003) outlined Success Based Evaluation models and there is much to recommend this method to study schools during their debates about National Standards. We have used this method in many evaluations in New Zealand on the Flaxmere Project, Gifted Kids, Healthy Housing, et al.; Clinton & Hattie, 2004, 2005).

One of the major topics should relate to the use of assessment information to inform the OTJs. As noted above with respect to the implementation of NCEA, there needs to be a realisation that the implementation issues are primarily about teachers' and students' *assessment capabilities* – NOT to do tests, but to use assessment in its myriad ways to make decisions about where we are going, how are we going, and where to next? Lester Flockton (2011) has presented on the DANZ (Absolum, Flockton, Hattie, Hipkins, & Reid, 2009) at this conference.

Let us not pretend how difficult it is for teachers to make and then defend their OTJs. Val Klenowski and colleagues (2009, 2011) have demonstrated this in their Queensland context, and we too need to see this as a worthwhile challenge – how to do this properly and validly – as then the students will be the beneficiaries. The recent NZCER review (Hipkins & Robertson, 2011) found little on how teachers learn together about moderation, although there are handbooks, special issues of journals, and a huge corpus of research on this topic when looking through the lens of human judgement decision making, moderation of teachers' interpretations of test scores, and there are well over 50 credible methods for standard setting and their effects on teachers. Until we learn to use this evidence then it is unlikely we will re-discover it, but then I suppose that is what re-search means!

If we do not get OTJs correct, then the march towards replacing teacher judgements with tests will be strong and hard to resist. Heaven help us if we have a NCLB, SATs, or NAPLAN. They are the hardest to get rid of, and a fear should be the profession defaulting to these seemingly easy options (with the nastiest backwash effects).

If we do not change teaching and learning, we change nothing worthwhile – provided the changes in teaching and learning are in the right direction. Where are the success case evaluations of how schools are using National Standards to enhance their teaching and learning, because if there is no change the policy will fail – but more important is to see how schools interpret policy to benefit students, or not.

Research Issue #6. The Impact on Students – Here Comes “Value Added”

It will not be long before hierarchical linear modelling (HLM) will be part of our research landscape. It is the fastest growing research method in the US and UK, and helps answer questions about value added, the effects of teachers and classrooms, students, and the interactions between the teachers and students. It assumes that students are nested in classes, nested in schools, and that students start at different points and this needs to be considered in any growth claims (Hodis, Meyer, McClure, Weir, & Walkey, 2011). Yes, like any technique it has been used in some disputed work, such as when the LA Times printed the value added for all LA teachers with the mistaken assumption that student value-added can be solely attributed to one teacher. We have a long history in measurement on care about causes and correlations but this does not stop some from being simplistic.

The problem in New Zealand is the lack of training in our universities in measurement and statistics. The New Zealand Assessment Academy (led by Gilmore) is one group trying to improve this training, but there are so few who research using HLM, IRT, SEM, and many other methods. There is a need for there to be experts on this method in New Zealand, before it is used simplistically and condemned as inadequate (when it is a perfectly acceptable method), and it can greatly aid in understanding how schools can enhance learning.

Research Issue #7. Evidence-based Teaching

There is also a great need for classroom observation research about what is different, or not, in classrooms or staffrooms as a consequence of this new policy. There are many Assessment for Learning (AfL) researchers in New Zealand, although the connection between National Standards and the importance of AfL seems too rarely made.

Evidence is a contested notion, as it should be. It could be simple as using test scores to make decisions; it could be using test scores to inform professional judgements; and it could be using other methods (like growth effect-sizes; Hattie, 2011) to inform debates among teachers and students. I have argued (Hattie, 2009) that there is a practice (note, I do not say science) of teaching – we know a lot that works best and does not work best; we have standards of acceptability (e.g., $d = .40$ is a norm growth “hinge point”); and we do have many methods to provide evidence to inform our educational debates.

The reactions to National Standards could well be monitored to evaluate the nature of ‘evidence’. Given what is happening in the policy world any move to not engage in this “evidence-based” movement will leave teaching well behind in convincability (and thus not lead to education gaining a decent percentage of the financial pie).

Conclusions

The research on National Standards in New Zealand is too little, and one hopes not too late. While we wail about similar models introduced overseas and their positive and negative effects, there are some key differences. On the one hand, the National Standards messages seem to be a road down the “testing is good” pathway; the centrality (if it remains) of OTJs places the onus of the policy on the teacher professional judgements – their interpretation of the test information (from multiple sources) and not on the scores from any one or a few tests. This is dramatically different from the US, Oz, and UK.

We have shown in New Zealand that poorly implemented policies can have markedly positive effects – just as I would claim that the NCEA has on high school teaching, on retention, and overall student success – notwithstanding that it has introduced problems and issues of its own.

This conference was conceived as a forum to see the effects of assessment practices in our schools – I suggest to you that its true success will be the ways researchers collaborate and cooperate in not only learning more about the impact of policies like National Standards and NCEA, but sharing their findings and implications across all who have a vested interest. Given the current research situation, I trust this conference will engender the participants to create a vibrant research plan to evaluate what might be our last chance before the introduction of national tests, private-public schooling, and the demise of teaching as a respected profession. Note, I am NOT arguing we must support the policy to research it. This is high stakes, and we can be an important part of it. Onwards a national research plan.

References

- Absolum, M., Flockton, L., Hattie, J. A. C., Hipkins, R., & Reid, I (2009). *Directions for assessment in New Zealand: Developing students' assessment capabilities*. Wellington, NZ: Ministry of Education. Retrieved from <http://assessment.tki.org.nz/Assessment-in-the-classroom/Directions-for-assessment-in-New-Zealand-DANZ-report>
- Baker, E. L., & Linn, R. L. (2004). Validity issues for accountability systems. In S. Fuhrman & R. Elmore (Eds.), *Redesigning accountability*. New York: Teachers College Press.
- Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, 17(2), 215-232.
- Brinkerhoff, R. O. (2003). *The success case method: Find out quickly what's working and what's not*. San Francisco: Berrett-Koehler.
- Clinton, J. M., & Hattie, J. A. (2004). *Evaluation of the Gifted Kids Program*. Auckland, NZ: Uniservices: University of Auckland/Ministry of Education.
- Clinton, J. M., & Hattie, J. A. C. (2005). *When the language of schooling enters the home: Evaluation of the Flaxmere Project*. Final report to Ministry of Education.
- Connolly S. R., Klenowski V., & Wyatt-Smith C. (2011). Moderation and consistency of teacher judgement: Teachers' views. *British Education Research Journal*, 38(4), 593-614.
- Cuban, L. (1993). *How teachers taught: Constancy and change in American classrooms, 1890-1990*. New York: Teachers College Press.
- Darling-Hammond, L. (2009). *The flat world and education: How American's commitment to equity will determine our future*. NY: Teachers College Press.
- Fiske, E. B., & Ladd, H. F. (2000). *When schools compete: A cautionary tale*. Washington, DC: Brookings Institution Press.
- Fletcher, R. B., & Hattie, J. A. C. (2011). *Intelligence and intelligence testing*. Oxford, UK: Routledge.
- Fullan, M. (2011). *Change leader: Learning to do what matters most*. Jossey-Bass.
- Hattie, J. (2007). The status of reading in New Zealand schools: The upper primary plateau problem (PPP³). *Reading Forum*, 22(2), 25-39.
- Hattie, J. A. C. (2009a). *Horizons and whirlpools: The well travelled pathway of national standards*. Retrieved from the Cognition Trust Website, www.cognition.co.nz
- Hattie, J. A. C. (2009b). Narrow the gap, fix the tail, or close the curves: The power of words. In C. Rubie & C. Rawlinson (Eds.), *Challenging thinking about teaching and learning*. Nova Science.
- Hattie, J. A. C. (2009c). Tomorrow's schools – Yesterday's news: The quest for a new metaphor. In J. Langley (Ed.), *Tomorrow's Schools: 20 years on* (pp. 121-134). Auckland, NZ: Cognition Institute.
- Hattie, J. A. C. (2009d). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Oxford, UK: Routledge.
- Hattie, J. A. C., & Peddie, R. (2003). School reports: "Praising with faint damns". *Set: Research information for teachers*, 3, 4-9.
- Hipkins, R., & Robertson, S. (2011). *Moderation and teacher learning: What can research tell us about their interrelationships?* Wellington, NZ: NZCER.
- Hodis, F. A., Meyer, L. H., McClure, J., Weir, K. F., & Walkey, F. H. (2011). A longitudinal investigation of motivation and secondary school achievement using growth mixture modeling. *Journal of Educational Psychology* 103(2), 312-323.
- Klenowski V. (2011). Assessment for learning in the accountability era: Queensland, Australia. *Studies in Educational Evaluation*, 37(1) 78-83.
- Koefed, W., et al., (2011). *A snapshot of written reporting practice and a review tool*. Wellington, NZ: Ministry of Education.
- Kofoed, W. (2009). *Written reporting: Strengthening learning partnerships through purposeful reporting* (Unpublished thesis). University of Auckland.

- Leeson, H. V., Brown, G. T. L., Hattie, J. A. C., Irving, S. E., Reddish, P., MacKay, A. J., & MacCall, S. (2005). *NZ students' performance in reading* (asTTle Monograph No. 3). Auckland, NZ: University of Auckland & Ministry of Education.
- Leeson, H. V., Hattie, J. A. C., Brown, G. T. L., Irving, S. E., Reddish, P., MacKay, A. J., & MacCall, S. (2005). *NZ students' performance in mathematics* (asTTle Monograph No. 2). Auckland, NZ: University of Auckland & Ministry of Education.
- Leeson, H. V., Hattie, J. A. C., Brown, G. T. L., Keegan, P. J., Irving, S. E., Reddish, P., ... & MacCall, S. (2005). *NZ students' attitudes to reading, writing, & mathematics* (asTTle Monograph No. 1). Auckland, NZ: University of Auckland & Ministry of Education.
- Levin, H., Belfield, C., Muennig, P., & Rouse, C. (2006). *The costs and benefits of an excellent education for all of America's children*. New York: Teachers College.
- Linn, R. L., Baker, E. L., & Betebenner, D. W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31(6), 3-6.
- Reddish, P., Brown, G. T. L., Hattie, J. A. C., Irving, S. E., MacKay, A. J., & MacCall, S. (2005). *NZ students' performance in writing* (asTTle Monograph No. 4). Auckland, NZ: University of Auckland & Ministry of Education.
- Rubie-Davies, C. M. (2007). Classroom interactions: Exploring the practices of high and low expectation teachers. *British Journal of Educational Psychology*, 77, 289-306.
- Rubie-Davies, C. M., Hattie, J. A. C., & Hamilton, R. J. (2006). Expecting the best for New Zealand students: Teacher expectations and academic outcomes. *British Journal of Educational Psychology*, 76, 429-444.
- Rubie-Davies, C. M., Hattie, J. A. C., Townsend, M. A. R., & Hamilton, R. J. (2007). Aiming high: Teachers and their students. In N. Galwey (Ed.), *Progress in educational psychology research* (pp. 65-91). Hauppauge, NY: Nova.
- Scriven, M. (1972). Pros and cons about goal-free evaluation: *American Journal of Evaluation February 1991*, 12(1) 55-62.
- Thrupp, M., Hattie, J., Crooks, T., & Flockton, L. (2009, November). *Open letter to the Minister of Education: Warning about the new National Standards system*. Retrieved from www.scoop.co.nz/stories/ED0911/S00107.htm

Researching Quality and Equity in Assessment for Improved Learner Outcomes

Val Klenowski
Queensland University of Technology

Email: val.klenowski@qut.edu.au

Abstract

Quality and equity issues as they relate to assessment practices and policy are becoming increasingly important as Australia introduces a national curriculum and achievement standards. In a context of high-stakes accountability, concerns regarding equity and quality have important implications for teachers' practice for the improvement of learner outcomes. This article is based on three research projects that were conducted over the past four years and were funded by the Australian Federal Government. The research focus, emergent questions, the educational contexts, and the rationale for the studies are discussed prior to the presentation of the analysis of the research findings and the implications for teachers' practice and policy reform.

Introduction

International comparative analyses of student achievement data, such as the Programme of International Student Assessment (PISA) developed by the Organisation for Economic Co-operation and Development (OECD) or the Trends in International Mathematics and Science Study of the International Association for the Evaluation of Educational Achievement (IEA), have influenced international curriculum and assessment reform agendas such that accountability, measurement and performance standards now dominate. This article draws on three recent studies conducted in this context of change where issues of equity and quality in assessment practices and policy, and the implications for learner outcomes, have formed the foci for the research.

The three projects, the emergent questions, the context, and the rationale for these studies will now be discussed followed by analyses of the research findings and the implications for teachers' practice and policy reform. Salient information pertaining to the three research projects, and the methodological and theoretical frameworks for each study, are provided to support the argument regarding the need to continue to theorise equity and related issues of quality in assessment to improve learner outcomes.

The Research Projects

The Australian Research Council (ARC) Linkage programme funded each of the research projects referred to in this article. This ARC programme requires the researcher to involve an industry partner who contributes funds and infrastructure support for the investigation. In each of these studies the partner investigators were interested in understanding how learner outcomes could be improved for all students, and how teacher assessment practice could be enhanced to address the important priorities of equity and quality in assessment practices.

The first study was concerned with the introduction of performance standards in Queensland where teacher assessment had been the dominant practice. The Queensland Studies Authority (QSA) was interested in providing appropriate support for teachers in their professional judgement practice of using achievement or performance standards for the first time. The QSA was willing to work with academic researchers to achieve improved understanding of the level of support required not just in terms of resources, but also conceptual tools, and informed process and policy that would be required to help

establish the conditions for improving student learning in this changed context of standards-referenced assessment. The project was entitled *Investigating standards-driven reform in assessment in the middle years of schooling*.

The second research project was deemed particularly timely and necessary against the background of Australia's underachievement in terms of equity for Indigenous students and the lack of an informed strategy in the education sector to counter this trend. The aims of the project were to explore and to develop an understanding of what culture-fair assessment might mean for Indigenous students. A major goal was to provide greater understanding about how to build more equitable assessment practices to address the issue of underperforming Aboriginal and Torres Strait Islander (ATSI) students in regional and remote Australia. This project was entitled *Developing culturally-fair assessment practices to achieve greater equity and success for Indigenous students* and the industry partners were the Catholic Diocese of Northern Queensland and the Association of Independent Schools, Queensland.

Sustainable selves: A new assessment model for marginalised secondary students was the third project that reconceptualised assessment as a form of exchange (Connolly, 2011). The aim was to explore how marginalised secondary students could use an Electronic Portfolio System (EPS) to document their own education development while providing the Industry Partner (Edmund Rice Education Australia) with a system to report and track young people's achievements.

Emergent Research Questions

Three key research questions emerged from these studies. These questions, which follow, raise important propositions for teachers' assessment and pedagogic practices, and for system level policy reform:

- What are the implications for teachers' assessment practice in the move to a standards-referenced system?
- How can culture-responsive assessment and pedagogic practice achieve fairer assessment?
- How can teachers increase participation for all students using an electronic portfolio system to support and assess achievement?

It is important to note that each study was concerned with equity, in terms of improving learning outcomes for *all* students. A major shift in emphasis occurred in the second study from a focus on *culture-fair* assessment to culture-responsive assessment and pedagogy. The sociocultural theoretical framing and approach adopted for the interpretative analysis of the data of this study helped the researchers identify how assessment could never be culture-fair. As expressed by Gipps and Murphy (1994), there is "[n]o such thing as a fair test and nor could there be as the situation is too complex and the notion too simplistic" (p. 273). We did, however, agree with Stobart (2008) in that "[w]e will never achieve fair assessment but we can make it fairer" (p. 113). Assessments have to be as fair as we can make them and issues of student access to curriculum and the framing of assessment tasks require teacher understanding of the relationships between teaching, curriculum and assessment, particularly in the increased high-stakes accountability context.

Global Context of Policy Reform

The rationale for these research projects that focus on quality and equity is derived from changing international and national contexts. Assessment and curriculum reforms have occurred largely due to the shifts in education policy processes. Fazel Rizvi and Bob Lingard (2010) make reference to Mahony, Hextall, and Menter (2004) who indicate that a

particular nation's education policy community of policy agents and agencies now extends to include global considerations. Rizvi and Lingard claim that the context of education policy production has changed to include a complex rescaling across local, regional, national, and global spheres. This new context for education policy production emphasises global competitiveness and global imperatives.

Global considerations are transforming "the balance between economic efficiency and the social equity goals of education" (Rizvi & Lingard, 2010, p. 15). The OECD policy framework that gives priority to market efficiency concerns at the expense of equity best illustrates this. Such economic reframing of education and education policy now emphasises production of human capital to ensure the competitiveness of the national economy in the global context (Rizvi & Lingard, 2010). The most evident example of global competitiveness is PISA which has had a huge influence on curriculum and assessment policy development worldwide. The media in any country is quick to emphasise international comparative analyses when PISA results are made available, and national governments respond by introducing change to national curriculum and assessment policies.

Australian Context of Educational Reform

In Australia, for example, the Australian Curriculum, Assessment and Reporting Authority (ACARA) has been established and is responsible for developing a national curriculum made up of content descriptions and achievement standards. Australian states and territories maintain the responsibility for assessment and reporting and are working with ACARA towards a more nationally consistent approach. The Australian curriculum makes the curriculum entitlement for all students explicit with the aim to address the issue of equity.

An Australian curriculum is in development with the following learning areas: English, Mathematics, Science, and History, being developed in Phase 1; Languages, Arts, Geography with Design in Phase 2; and finally Technology, Health and Physical Education, Information and Communication Technology (ICT), Economics, Business and Civics and Citizenship in Phase 3. In addition to the Learning Areas it is expected that the seven general capabilities listed below, which comprise skills, behaviours and dispositions that students develop and apply to content knowledge and that support them in becoming successful learners, confident and creative individuals, and active and informed citizens, will also be developed. Throughout their schooling students use these capabilities in their learning across the curriculum, in co-curricular programmes and in their lives outside school. They are:

- Literacy
- Numeracy
- Information and Communication Technology (ICT) competence
- Critical and creative thinking
- Ethical behaviour
- Personal and social competence
- Intercultural understanding.

There are also the three cross-curriculum priorities that are to be embedded in all learning areas. The first of these addresses *Aboriginal and Torres Strait Islander (ATSI) histories and cultures* to allow all young Australians the opportunity to gain a deeper understanding and appreciation of ATSI histories and cultures, their significance for Australia, and the impact these have had, and continue to have, on our world. The second relates to *Asia, and Australia's engagement with Asia*, to allow all young Australians to develop a better understanding of the countries and cultures of the Asia region. Students will be expected to develop an appreciation of the economic, political and cultural interconnections that

Australia has with the region. The third cross-curriculum priority is *sustainability* to allow all young Australians to develop an appreciation of the need for more sustainable patterns of living, and to build the capacities for thinking and acting that are necessary to create a more sustainable future. It is anticipated that these cross-curriculum priorities will have a strong but varying presence in the Learning Areas depending on their relevance.

In terms of assessment, the following types are practised in Australia. Students receive an A to E report card every semester in every year in every subject. School-based assessment comprises the National Assessment Program Literacy And Numeracy (NAPLAN) tests conducted in May – students receive their results in September, and the school receives diagnostic information in December or January. Parent assessment also takes place through observation and after-school study.

In Queensland, the QSA has developed the Queensland Comparable Assessment Tasks that are conducted in Years 4, 6 and 9. These tasks are designed to assist teachers to understand the qualities needed in student work indicative of the achievement standards. The tasks are considered to be “rich tasks” and are intended to demonstrate to teachers how they can design assessment to provide students with the opportunity to demonstrate understanding, as well as higher order skills, such as reflection and critical thinking. Other forms of testing at the school level are localised and include classroom testing and diagnostic testing with the use of tests such as the Progressive Achievement Tests in Reading developed by the Australian Council for Educational Research.

In 2010, the high-stakes nature of NAPLAN testing was confirmed with the publication of results on the MySchool website (www.myschool.edu.au/). The Federal Government reported high levels of parental support for this initiative and claimed that the website serves the best interests of transparency and accountability, it enables parents to have access to evidence of schools’ performance in NAPLAN testing, and that these published results support parental access to information about the quality of schooling, when in fact what has occurred is that the results themselves have become codes or indexes for the quality status of individual schools and education systems more generally. The suggestion that parents would have the necessary data for choice of school for their children was made, yet very little was stated about the issue of equity.

The Australian government’s push for high-stakes testing is driven by a desire to meet public accountability, demonstrate transparency, and maintain public confidence in the standards of schooling. The issue for educators across Australia is whether this push will meet accountability demands without negatively impacting on the quality and equity of teaching and learning.

Assessment Reform and Unintended Consequences

Accountability testing has assumed increasing prominence in public education policy internationally. With this move towards more summative and high-stakes testing there have emerged important messages for policy developers and practitioners about unintended consequences. In the USA, research indicates that when there is a shift to focus only on the results in the evaluation of school performance and when sanctions are imposed, negative consequences follow. Nichols and Berliner (2010) in the US have researched how increased pressures on teachers to ensure improved tests results lead to cheating. They studied the types and degrees to which a sample of teachers were aware of, or had participated in, these practices because of the *No Child Left Behind* high-stakes testing policies. They concluded that:

Policies that clearly undermine the moral and professional behaviour of America’s teachers need to be debated more thoroughly, and such policies must be challenged if their negative effects outweigh their positive effects on the educational system of our nation. It serves no one’s interest to have policies that inherently promote cheating, and even justifications for cheating by educators, because the policy

environment in which they work has become so onerous. There are better ways to design accountability systems. (p. 27)

In Australia, 2008 marked the extension of Australia's NAPLAN testing from Years 3, 5 and 7 to include students in Year 9, and a concurrent move driven by Commonwealth funding legislation, from state-based to a national testing system. From 2006, schools were required to make their individual school performance test data (literacy and numeracy) available publicly. It is important to reflect on the direction Australia is taking in terms of high-stakes assessment programmes such as NAPLAN because it is this type of programme that can have unintended and negative effects on teaching and learning quality. In Australia, examples of cheating have emerged and have been reported in the public media. There have also been pressures on leaders to lift performance, threats to their jobs if results do not improve, more attention given to those students who are likely to achieve better grades, neglect of those students who have the greatest need for support, the emergence of commercial tests that have not been quality assured, and increased absenteeism for low performing students on the day of the test (Klenowski & Wyatt-Smith, 2011).

My argument is that schools must be protected from such consequences. The way forward in contexts of major assessment and curriculum reform is policy informed by research to support and manage issues of quality and equity. The findings from the three research projects indicate that teachers' skills and understandings about inclusive and ethical practice are priorities for support. Teaching and assessment that is responsive to sociocultural contexts and cultural and social difference help to address questions of equity. Alan Reid (2011) has raised the important point in this changing educational context that only when efforts are made to theorise equity as a concept and a practice will policy related to equity have an impact.

Key Messages

The key messages from the research and the literature are that it is important how results are interpreted and presented. Raw scores are not helpful because related equity issues are not made explicit. Too often there is over-interpretation of students' results in terms of innate ability, dispositions, and limitations (Klenowski, 2009; Murphy, 2009). 'High expectations' for all students and 'effortful teaching' (Klenowski, Tobias, Funnell, Vance, & Kaesehagen, 2010) become priorities.

The key findings from the research and the literature to date are that:

- research-informed policy is needed to prevent unintended consequences (Klenowski & Wyatt-Smith, 2011);
- equity needs to be theorised as a concept and a practice (Reid, 2011);
- participation in learning and assessment requires effortful teaching for all students to be involved;
- teachers require support, particularly in relation to assessment literacy (Hipwell & Klenowski, 2011); and
- standards, and assessment in general, can be used to support learning (Wyatt-Smith & Klenowski, 2010).

The most central and important change agent is the teacher and the quality of the teacher-student relationship in the community of learning that exists in the classroom. My argument is that to improve the quality of assessment practice it is important to provide opportunities for teachers to develop their assessment literacy (Hipwell & Klenowski, 2011). Resources and scheduled time need to be provided for teachers to meet to moderate for dependable judgments. Specific support for teacher classroom assessment to directly contribute to improved learning for all students is also a priority. Teacher

development requires theoretically-based approaches yet practically-situated learning for both pre-service and in-service teachers to build their assessment repertoires based on the development of their own understandings of theories of learning and assessment.

Assessment and Learning

A sociocultural theoretical stance was adopted in each research project to explain and identify the implications for building teachers' assessment repertoires and assessment identities. From this theoretical position, learning is understood as "becoming" a full participant in the community of learners whether that is a teacher becoming more competent in making judgments that accurately reflect the standard of student work, through interaction with others and the standards in moderation practice, or whether it is a student whose identity as a learner is strengthened through increased participation and re-engagement with learning as a consequence of increased interactions and participation with the community of learners in the mathematics classroom, or via the internet using digital means. The sociocultural view sees learning as a sense of belonging through increased participation and change of identity to acceptance through effortful teaching and improved relationships.

Messages from the Research on Standards-based Assessment

Social moderation processes are necessary in a standards-based assessment system as teachers' school-based assessment is usually deemed to have high validity but questionable reliability. The defined standards can inform teacher judgement of system level expectations (Klenowski & Wyatt-Smith, 2010); however, a standards-based assessment system requires teachers to be skilled in assessment task design and to appreciate the importance of the relationship between curriculum, assessment and pedagogy. Teachers' involvement of students as participants in the use of standards for self and peer-assessment to learn about the qualities indicative of the standards is expected from this theoretical stance.

The findings from this project identified how teachers' participation in moderation practice supports their judgements of the standards of the student work and that this practice helps to achieve fair and dependable judgements (Wyatt-Smith, Klenowski, & Gunn, 2010). Moderation also can address equity concerns in that the practice is responsive to a wide range of evidence types and contexts. However, to support moderation practice a suite of resources is needed – guides to making judgements, a statement of the standards, illustrative exemplars of work, folios or body of evidence, and a cognitive commentary that explains the trade-offs and/or compensatory factors in the processes of arriving at an overall judgement.

Moderation is conceptualised as a social practice of interaction and exchange of views of quality of student work with reference to the standards for achieving inter-rater-reliability and more comparable and consistent judgement. In this context, standards are intended to be used as the basis for judgements of student achievement, while the results from assessment tasks are meant to both inform the teaching/learning process, and to report and track student progress. In such a system, the role and reliability of teacher judgement takes centre stage.

Social moderation provides a learning opportunity for teachers in the generation of new knowledge and new ways of knowing. The teachers in the moderation meetings discuss and debate their interpretations of the quality of the evidence in terms of the award of the particular achievement standard. This is a context of situated interaction where teachers participate drawing on their individual tacit and explicit knowledge of standards together with the group's explicit and tacit knowledge of the standards. The source of new knowledge and knowing lies in the interaction of all knowledge (tacit and explicit) about

standards as a tool of knowing or as described by Cook and Brown (1999)¹ in 'the generative dance' or knowing in action.

The purpose of social moderation is to produce valid and reliable judgements that are consistent with one another and with stated standards of performance. It is a form of quality assurance in terms of achieving comparability of judgements based on evidence of student achievement (Maxwell, 2007). Moderation is also a social practice that encourages the development of a sense of community of assessors.

Teachers' engagement in moderation practice and the new knowledge and ways of knowing that are generated are as follows:

- Teachers are able to check that similar skills and levels of skills are taught and similar outcomes are assessed as equitable and of a comparable quality.
- Fairness for all students is extended beyond the classroom or school to between schools spread within the state.
- Increased confidence for teachers, parents, students, and other staff members that common standards are expected and are being achieved by a particular year group of students
- Teaching and assessment practices are made transparent. Teachers' work is made public, and open to scrutiny and critique that helps to address accountability and quality assurance demands.
- Gaps or omissions in the teaching programme can be identified, particularly if the Director of Curriculum or Head of Department participates in the moderation meetings.
- A sense of community develops as teachers negotiate their understanding and seek clarification and advice when they are unsure of the standard or the standard of work. There is a shift from individual practice to shared practice and the improvement of shared practice.
- Engagement in moderation practice focuses teachers' attention on assessment and its place within the teaching and learning programme. Teachers seem inspired to teach a topic when they realise the results achieved by other teachers using different approaches. In this way teachers learn new ways of teaching a topic, are diversifying their practice to meet the needs of individuals and in so doing are improving practice (Klenowski & Adie, 2009).

Internationally moderation has been recognised as an important practice with the Cabinet Secretary for Education and Lifelong Learning in Scotland stating that:

A national system of quality assurance and moderation of 3-18 will be developed to support teachers in achieving greater consistency and confidence in their professional judgements. (Hyslop, 2009)

In Wales, national tests have been abandoned and the value of school-based assessment and teacher moderation practice has been recognised. Recommendations from the Report on *Future assessment arrangements for Key Stages (KS) 2 and 3* included cluster group moderation for transition links with Key Stage 2 and 3 schools. Strengthened assessment at the end of KS 3 takes place by means of external moderation of sample evidence of teachers' understanding and application of the national curriculum level descriptions.

A major challenge for the current curriculum-driven reform in Australia is for the relationship between the learner, learning, and assessment to remain central and for the professionalism of teachers to be sustained through more helpful approaches to assessment beyond national tests.

¹ I am indebted to Patricia Murphy for drawing this concept to my attention.

Fairness in Assessment

In this context of major assessment and curriculum reform, policy and research to support and manage issues of quality and equity are helpful to up-skill teachers in inclusive and ethical assessment practice. A shift to focus on learning for *all* students can occur with change in teaching and assessment such that it is more responsive to cultural and social difference. An aim of the research reported here was for educators to take care in how the achievement results were interpreted and presented, seeing beyond the raw scores to understand the related equity issues with no over-interpretation of students' results in terms of innate ability and limitations. The move in the second project from a focus on culture-fair to culture-responsive or culture-aware assessment was demonstrative of this intended aim. Access issues and the literacy demands of the assessment tasks and tests progressively became the priorities for study. The opportunity for all students to demonstrate their learning was another intended outcome in relation to the validity of the tasks and assessments designed and developed at local and state levels, for as expressed by Gipps (1999), "Openness about design, constructs and scoring, will bring out into the open the values and biases of the test design process, offer an opportunity for debate about cultural and social influences, and open up the relationship between the assessor and the learner" (p. 385).

In this study it became apparent that many teachers were unaware that there is no cultural neutrality in assessment or in the selection of what is assessed (Cumming, 2000). "When setting standards and test content, are we really sure that this is the knowledge we need?" or "Are we really privileging certain knowledges to maintain a dominant culture, and, in so doing, ensuring perpetuation of ourselves, as people who have succeeded in the formal educational culture to date?" (Cumming, 2000, p. 4). The project sought to raise teachers' awareness of group differences, interaction of mode of assessment with construct assessed and student experience, society's views and expectations of the abilities of different student groups, and how this impacts on expectation, teaching and curriculum offered (Gipps & Murphy, 1994).

Teachers were informed that students from a non-dominant culture can experience testing and assessment as a form of cultural intimidation and may even develop resistance to the surveillance, judgement and categorisation practices that are affiliated with large-scale testing (Berlack, 2001). It was demonstrated to the teachers how cultural differences can impact on student performance in the context of standardised tests such as NAPLAN and how equity issues relate to opportunity to participate and to demonstrate learning. This was achieved through fine-grained analyses of Indigenous students' responses to each item on the test together with the teachers' understanding of the students' disposition and performance in school-based assessment tasks.

Fairness of any test or assessment depends on whether the students are able to make sense of what is required. It is important for those students from culturally and linguistically diverse groups or those from backgrounds of poverty and social disadvantage to be provided with opportunities to offer alternative evidence of their expertise. To achieve assessment that is as fair as possible teachers use a range of modes and task styles. The use of rich tasks and more authentic assessments was an attempt in the culture-responsive assessment project to achieve this goal.

As educators and researchers, it is my argument that we need to raise teachers' awareness of group differences and ensure that we have high expectations for all students. We need teachers to understand the interaction of the mode of assessment with the construct assessed and student experience to ensure that there is access for all in terms of what is being asked in the assessment task or test. In teaching mathematics we found it is important to understand the students' understandings, dispositions, self-beliefs and acknowledge their personal view of the value of learning (mathematics). This required the teacher to use assessment formatively to interact with the student to help the student re-imagine a different identity as a successful learner (of mathematics).

We found that the majority of teachers interviewed had no professional development in relation to Indigenous cultural awareness and that the language issues and literacy demands of the assessments and tests were not always addressed in pedagogic practice. The students' sociocultural circumstances needed to be understood by the teachers and school leaders as students' attitudes to learning are directly affected by the value they place on the learning and the success that they believe they might have in reaching a satisfactory goal. We concluded that strategic and effortful teaching (Klenowski et al., 2010) that encompasses a diagnostic and holistic view of a student's background, culture, language and demeanour was central for developing mathematical thinking. We also raised the issue of how schools could develop their capacity to identify 'deficit views of difference' (Ainscow, 2009) whereby students are seen as 'lacking in something'. We aimed to challenge teachers' assumptions related to notions of deficit regarding difference using a sociocultural perspective of learning and assessment. From this theoretical position there is greater respect for valuing of difference.

Assessment as Exchange

The focus of this third project was on the improvement of learner outcomes for those students who had left formal education and were seeking to re-engage through alternative programmes such as those offered by the Queensland Government's Education and Training Reforms for the Future programme. The aims of this programme are to address questions of equity and opportunity by seeking to re-engage students and to provide for them through socially just or altruistic means. This particular study was conducted in one of the largest re-entry programmes, the Flexible Learning Centre Network. The focus was on improving educational achievements by introducing a new model for the assessment of the progress of these young people. Authentic assessment and assessment for learning were key elements of the model designed to develop and implement individual portfolios. Based on sociological models of capital (Bourdieu, 1986), these portfolios were compiled using quantitative and qualitative evidence of young people's resources and achievements. Further aims were: to enable young people to understand and document their own education development; to provide teachers and para-professionals with new grounds for curriculum and counselling; and to provide systems of reporting and tracking of young people's development for funding and accountability purposes.

Assessment in this context is conceptualised as a form of exchange and is enacted using an EPS (Connolly, 2011). This new conceptual model and online support tool for assessing the educational progress of disengaged youth who re-enter education via a specific network of Flexible Learning Centres (FLC) in Queensland, Australia was the focus for the research. The FLC network provides a blend of formal education: nationally certified courses and formal curriculum material with social, cultural and arts-based education support projects that are informal in their delivery and assessment.

From a sociocultural view of assessment for learning teachers are positioned as learners in a highly communicative context. The communication takes place in the field of education where agents exchange cultural capital (Bourdieu, 1986). Within the field of education, agents exchange cultural capital through the use of an assessment for learning approach enacted largely on the EPS, a hybridised electronic portfolio, and a Content Management System (CMS), two online systems operating as one. It was hosted at Queensland University of Technology and provided a high level of communication diversity and unique assessment experiences (Connolly, 2011).

The term the 'efield' (Connolly, 2011) was coined to describe the social networking and content management system that constitutes the learning and assessment environment. An assessment for learning approach was adopted to provide the opportunity for these young people to draw on the social and cultural capital they bring to the assessment interaction or exchange. For some this is a musical talent, for some it is an artistic endeavour such as photography or drawing, and for others it is an interest in a particular

field such as surfing or cycling. Using the workspace the young people are able to exhibit their portfolios of work and can self and peer assess this body of evidence.

In designing assessment tasks and processes to engage or re-engage students, opening up realistic opportunities to participate becomes a priority. Multiple indicators of achievement and opportunities to offer alternative evidence of expertise are clear messages to policy developers to support quality education for all students while meeting accountability demands. It is also helpful to consider the relationship between access, curriculum, and assessment and to ask these important equity questions.

Table 1: Curriculum, Assessment and Equity Questions (Stobart, 2005, 2008)

Access Questions	Curricular Questions	Assessment Questions
Who gets taught and by whom?	Whose knowledge is taught?	What knowledge is assessed and equated with achievement?
Are there differences in the resources available for different groups?	Why is it taught in a particular way to this particular group?	Are the form, content and mode of assessment appropriate for different groups and individuals?
What is incorporated from the cultures of those attending?	How do we enable the histories and cultures of people of colour, and of women, to be taught in responsible and responsive ways?	Is this range of cultural knowledge reflected in definitions of achievement? How does cultural knowledge mediate individuals' responses to assessment in ways which alter the construct being assessed?
(Stobart, 2005)	(Apple, 1989)	(Gipps & Murphy, 1994)

Conclusion

As researchers we need to be aware of the changing policy contexts at the global and national levels and we need to conduct research that will work to inform both practice and policy development. We need to generate more theorised understandings of how assessment can develop learning for all, yet engage with the priorities of accountability and the current educational priorities in a world of continuing change.

References

- Ainscow, M. (2009). Local solutions for local contexts: The development of more inclusive education systems. In R. Alenkaer (Ed.), *Den Inkluderende Skole I et Ledelsesperspektiv (Leadership perspectives on the inclusive school)*. Copenhagen DK: Frydenlund.
- Berlack, H. (2001). Race and the achievement gap. *Rethinking Schools Online*, 15(4). Retrieved from http://www.rethinkingschools.org/archive/_04/Race154.shtml
- Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241-258). New York: Greenwood.
- Connolly, S. (2011, August). *Towards general capabilities through technology: Assessment for learning as a field of exchange*. Paper presented at ACACA conference, Brisbane.
- Cook, S. D. N., & Brown, J. S. (1999). Bridging epistemologies: The generative dance between organizational knowledge and organizational knowing, *Organization Science*, 10(4), 381-400.

- Cumming, J. (2000). After DIF, what culture remains? 26th IAEA Conference, Jerusalem.
- Gipps, C. (1999). Socio-cultural aspects of assessment. *Review of Research in Education*, 24, 355-392.
- Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Buckingham, UK: Open University Press.
- Hipwell, P., & Klenowski, V. (2011). A case for addressing the literacy demands of student assessment. *Australian Journal of Language and Literacy*, 34(2), 127-146.
- Hyslop, F. (2009). *Assessment for curriculum for excellence*. Retrieved from http://www.ltsotland.org.uk/Images/AssessmentforCfE_tcm4-565505.pdf
- Klenowski, V. (2009, November). *Raising the stakes: The challenges for teacher assessment*. Keynote address, Australian Association of Research in Education Conference, Canberra, ACT.
- Klenowski, V., & Adie, L. E. (2009). Moderation as judgement practice: Reconciling system level accountability and local level practice. *Curriculum Perspectives*, 29(1), 10-28.
- Klenowski, V., Tobias, S., Funnell, B., Vance, F., & Kaesehagen, C. (2010, November). *Culture-fair assessment: Challenging indigenous students through effortful mathematics teaching*. Presentation at the Australian Association of Research in Education Conference, Melbourne.
- Klenowski, V., & Wyatt-Smith, C. M. (2010). Standards, teacher judgement and moderation in the contexts of national curriculum and assessment reform. *Assessment Matters*, 2, 107-131.
- Klenowski, V., & Wyatt-Smith, C. M. (2011). The impact of high stakes testing: The Australian story. *Assessment in Education: Principles, Policy and Practice*. Retrieved from <http://eprints.qut.edu.au/47882/>
- Mahony, P., Hextall, I., & Menter, I. (2004). Building dams in Jordan, assessing teachers in England: A case in edu-business. *Globalisation, Societies and Education*, 2(2), 277-296.
- Maxwell, G. S. (2007). *Implications for moderation of proposed changes to senior secondary school syllabuses*. (Paper commissioned by the Queensland Studies Authority). Brisbane: Queensland Studies Authority.
- Murphy, P. (2009, September). *Applying a sociocultural approach to assessment theory and practice: Issues for summative and formative assessment*. The Open University, Presentation at Queensland University of Technology, Brisbane.
- Nichols, S. L., & Berliner, D. C. (2010). *The inevitable corruption of indicators and educators through high-stakes testing*. East Lansing, MI: The Great Lakes Center for Education Research & Practice
- Reid, A. (2011). What sort of equity? *Professional Educator*, 10(4), 3-4.
- Rizvi, F., & Lingard, B. (2010). *Globalizing education policy*. London: Routledge.
- Stobart, G. (2005). Fairness in multicultural assessment systems. *Assessment in Education: Principles, Policy and Practice*, 12(3), 275-287.
- Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. London: Routledge.
- Wyatt-Smith, C. M., & Klenowski, V. (2010). Role and purpose of standards in the context of national curriculum and assessment reform for accountability, improvement and equity in student learning. *Curriculum Perspectives*, 30(3), 37-47.
- Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy and Practice*, 17(1), 59-75.

The NCEA as an Exemplar of Research into Practice: Motivation, Achievement and Assessment Design¹

Luanna H. Meyer
Victoria University of Wellington

Email: luanna.meyer@vuw.ac.nz

Abstract

This overview of a series of studies carried out by an interdisciplinary team of researchers at Victoria University highlights the impact of assessment policy and practice on achievement behaviours and actual educational outcomes for New Zealand secondary students. Mixed methods research approaches were used to investigate longitudinal influences on student motivation and achievement as a function of the National Certificate of Educational Achievement (NCEA). Highlights from survey, interview and achievement data are summarised with implications for assessment design, assessment practices, and strategies that teachers in particular can use to motivate students to do their best in their final years of secondary school.

Introduction

In New Zealand as in other nations, there is a need for educational research that tackles significant educational challenges and generates knowledge leading to enhanced educational outcomes for children and youth. Indeed, there should be widespread support for research that investigates the empirical base for national policy and for recommended best practices in our schools. It would also be helpful if our research agenda included development of do-able practices that can be carried out in typical classrooms and educational settings – going beyond carefully controlled research designs to demonstrate feasibility in the context of real schools with the adults and young people who populate our schools.

The National Certificate of Educational Achievement (NCEA) represented a major shift in assessment policy and practice for this country – thus also providing an opportunity to examine empirically the validity of a new system designed to better respond to educational imperatives in today's world. This paper highlights key findings from a series of research studies encompassing:

- Large-scale research conducted in nationally representative secondary schools emphasising student voice
- Advanced mixed methods approaches that support the generalisability (quantitative) of findings and the credibility (qualitative) of interpretations about those findings
- Empirical investigation of major theories of achievement motivation and planned behaviour
- Interdisciplinary perspectives and strengths (education and psychology)
- Longitudinal tracking of multiple cohorts of students across time and as a function of major NCEA design changes that might be expected to have an impact on student motivation and achievement.

Issues Emerging from the Research

The research team carrying out this work has to date published several reports on the MOE Education Counts website as well as a series of journal articles describing major findings so this keynote is not meant to repeat this work. Instead, issues reflecting three major themes are emphasised here: (1) the impact of assessment design on motivation

and achievement; (2) predicting in order to prevent underachievement across students; and (3) the identification of strategies teachers and other educationalists can use in schools and classrooms to enhance student motivation and achievement.

Theme 1: Impact of Assessment Design on Motivation and Achievement

We have identified several motivation orientation factors associated with secondary student achievement on the NCEA (Meyer, McClure, Walkey, Weir, & McKenzie, 2009). Students complete a brief survey (it can be completed during a typical Form period in secondary school) consisting of various statements that, taken together, have been factor analysed to identify meaningful dimensions of motivation associated with achievement and that may be amenable to change as a function of school strategies.

Two of the survey factors in particular are highly predictive of achievement 3-4 years later—*Doing My Best* and *Doing Just Enough* (Hodis, Meyer, McClure, Weir, & Walkey, 2011). These motivation orientations are listed here with the three statements for rating using a Likert scale to identify the extent to which each student considers that the statement is true for him or her.

Doing Just Enough

- Once I have my 80 credits, I'll be satisfied.
- If I get just NCEA Level 1 or possibly NCEA Level 2 before I leave school, I'll be satisfied and have no plans to finish Level 3.
- I will work for the number of credits I need at each level, no more.

Doing My Best

- I want to take credits that allow me to try for Merit or Excellence, rather than just Achieved.
- I will strive for Merit or Excellence even when I don't need this to achieve my goals.
- I expect to get Excellence or at least Merit when I do NCEA.

These are the kinds of statements that parents and teachers reported hearing from young people, suggesting implications for policy and school practices. If, for example, students were able to complete 80 NCEA achievement standard credits at Level 1 entirely through internal assessment and even early in the year, they would have little motivation to continue working through assessments beyond that point and sit external examinations at the end of the year. Given that an individual student is likely to have access to upwards of 120 credits in any given school year, our system could be inadvertently de-motivating students from bothering with a large proportion of learning assessment opportunities. Indeed, high scores on *Doing Just Enough* predicted fewer credits, as did low scores on *Doing My Best*. Our research team thus presented evidence that:

- Low aspirations predict low achievement, and high aspirations predict high achievement
- Low expectations for students reinforce low aspirations, which in turn would be associated with underachievement (at all levels, including for higher achieving students as well as those who may be struggling)
- Assessment design can result in either encouragement to aim high(er) to reach high(er) goals, or it can reinforce 'doing just enough'
- Assessment practices such as not recording 'not achieved' grades (as during early implementation of NCEA) and lock-step enrolment into unit standards (that did not offer any opportunities to work towards Merit and Excellence) will deny educational opportunities for enhancement, reinforce poor motivation and perpetuate low(er) achievement.

Consequently, we recommended and promoted the following changes in NCEA assessment policies and practices from 2006 to the present (Meyer, Weir, McClure, Walkey, & McKenzie, 2009):

1. Not Achieved: This grade is now included on a student's Record of Learning, along with the other three grades of Achieved, Merit and Excellence.
2. Unit Standards: These standards should be reviewed for conversion to achievement standards that would present opportunities to earn Merit and Excellence rather than only either Not Achieved or Achieved. We argued that surely all educational performance is worthy of recognition when performed at a high level, rather than seeing high grades as somehow relevant only to traditional 'academic' subjects. The availability of high grades across all subject areas could enhance motivation for a broader range of students.
3. Certificate Endorsements: Individual achievement standard credits could be endorsed for Merit and Excellence and these grades would appear on the record of learning. We argued, however, that students would be motivated to work even harder if they could earn Merit and Excellence across the entire NCEA Level Certificate.
4. Course Endorsements: Similarly, students could be motivated to work towards Merit and Excellence endorsements in individual subject/course areas. This could motivate both high achievers (who would want to get as many course endorsements as possible, across different subjects) and those students who may not be outstanding 'all-rounders' but have exceptional talent and interest in particular subjects (an equity issue again).
5. NCEA Level Attainment: We provided evidence that students who indicated they intended to attain only NCEA Level 1 or Level 2 were performing educationally much as they said they would.

Across our data, low aspirations were related to low achievement. We argued that students should be discouraged from low aspirations – that teachers and other school professionals should communicate high expectations and promote high aspirations. In particular, if students were allowed to take two school years – years 11 and 12 – to attain just NCEA Level 1, they were unlikely to reach even NCEA Level 2 prior to leaving school. This situation wasn't good enough!

During the first few years of implementation of the NCEA, with support from our research and educational advocacy from across the sectors, refinements and design changes were made to the NCEA that addressed each of these and we have been able to show (empirically) consequent student improvements in achievement.

Theme 2: Predictions Towards Prevention of Underachievement

Our survey is brief and could be administered and scored directly by any school in order to investigate student aspirations about their own achievement. Ideally, teachers and schools could use this information – gathered as early as Years 9-10 – to identify students who need a *motivation boost*.

Motivation messages can be quite simple and subtle. A common statement parents and teachers sometimes make is to encourage a young person to try something new and/or work on something, stating something like "You can do this – it's easy!" Telling a student that something is easy can devalue performance when he/she succeeds. It will also be devastating if the task doesn't feel easy to that student and/or if the student actually fails the task – leading to self-interpretations such as "If this is easy and I cannot do it, I am really thick." Hence, a teacher or parent should instead always say something like: "This is difficult and you'll have to work on it, but you can do it if you do work on it and I'll give you help if you need it" so that when the student does succeed, it feels like a real accomplishment.

Early in our research, we maintained that merit and excellence were relevant to work across the curriculum. Think about it: would you hire a builder who was *Doing Just Enough* or would you hire a tradesperson committed to *Doing My Best*? Why should merit and excellence be confined only to traditional academic subjects or only be relevant to students planning university study in the future? Some general messages from the educational system and the school overall can actually reinforce low achievement: for example, students who are tracked into unit standards that do not offer even opportunities to earn Merit are receiving strong messages that *excellence is not for you*. If every student in school was reminded by his or her teachers regularly that he or she can strive for credits attained with Merit or with Excellence as well as Achieved, this would be a strong positive message that we have high expectations for every student – and ultimately for graduates of our educational system regardless of their career pathway!

On our student survey, we ask students *What is the highest NCEA level you expect to complete before you leave school?* Three years ago when we asked this question of nearly 2,500 Year 11 students nationally, we received the following answers:

None:	N = 163 students
Level 1:	N = 420 students
Level 2:	N = 858 students
Level 3:	N = 3,848 students

Not surprisingly, students who reported they did not intend to complete any level of the NCEA expressed poor motivation patterns on the other sections of the student survey. However, students who indicated they intended to complete only Level 1 also expressed relatively poor motivation orientation patterns. Students indicating they intended to complete Level 2 or Level 3 did not differ from one another in educational aspirations. We maintain that our evidence on the relationship between these expressed aspirations and overall motivation orientations as well as student achievement patterns supports the need to encourage high aspirations for *all students*, with both teachers and students focused on NCEA Level 3 for everyone then determining how to accomplish that.

Elsewhere in these proceedings, Sanders (2012) reports on an intervention study in which a typical secondary school homework programme incorporated motivation-enhanced messages. This reports a component of his PhD research which demonstrates that teachers can change motivation messages and mindsets for students to positive effect so that student achievement is then enhanced accordingly. We need more work such as this!

Theme 3: Teacher Strategies to Enhance Motivation and Achievement

Another dimension revealed through factor analysis on our student survey also correlated with motivation and achievement – student ratings on items addressing *Teacher Affiliation* factors:

- I do best when the teacher expects me to do well.
- I do best when I know the teacher will help when I need it.
- I learn more in a subject when the teacher cares how well I do.
- I like a subject more when the teacher encourages me.

Our findings from the NCEA research highlight messages that teachers can promote actively to support strong motivation and higher achievement (Graham, Meyer, McKenzie, McClure, Weir, & Walkey, 2010; McClure, Meyer, Garisch, Fischer, Weir, & Walkey, 2011). These findings also align well with findings reported from implementation of Bishop's Te Kotahitanga research to enhance teacher capacity to promote high achievement for Māori students through culturally responsive pedagogies (Bishop, 2011; Hindle, Savage, Meyer, Sleeter, Hynds, & Penetito, 2011; Hynds, Sleeter, Hindle, Savage, Penetito, & Meyer, 2011; Penetito, Hindle, Hynds, Savage, & Kus, 2011; Savage, Hindle, Meyer, Hynds,

Penetito, & Sleeter, 2011). Clearly, there is much the teacher can do and does do in the day to day learning interactions that matters to students – and to how hard they try and how well they do.

Discussion: School and Teacher Influences on Outcomes

Can we predict and prevent underachievement—rather than wait for students to fail and then intervene? We have demonstrated that a simple screening tool that can be completed by students in less than 10 minutes is highly predictive of student achievement 1-3 years later (Hodis et al., 2011). Thus, we are able now to identify students as early as year 10 who candidly tell us what they intend to do in school, but the next step would be for us to intervene effectively to raise aspirations (theirs) and expectations (ours) where they are low. A self-rating such as this could also be done in Year 9, provided students have some understanding about the NCEA – which would seem important regardless. The important next step is not to reinforce those intentions or motivations, but to intervene through school and teacher practices that are feasible and effective (e.g., Sanders, 2012).

One of the positive features of the NCEA is the fact that it presents multiple opportunities for students to share information about their learning as they complete many different standards and work towards course and certificate achievements. There is, however, much more we can do and that still needs doing to take full advantage of this standards-based assessment system. For example, streaming students into unit standards coursework communicates powerful negative messages that we have low expectations for them so we should not be surprised if they have low aspirations. There are still additional design changes needed to address motivational challenges. And perhaps most important of all, we need systematic intervention research to investigate the kinds of approaches that work not just in principle or theory but make a real difference to educational outcomes for our young people.

References

- Bishop, R. (2011). Te Kotahitanga: Kaupapa Māori in mainstream classrooms. In C. E. Sleeter (Ed.), *Professional development for culturally responsive and relationship-based pedagogy* (pp. 23-45). New York: Peter Lang.
- Graham, J., Meyer, L. H., McKenzie, L., McClure, J., Weir, K. F., & Walkey, F. (2010). Māori and Pacific secondary student and parent perspectives on achievement, motivation and the NCEA. *Assessment Matters*, 2, 132-157.
- Hindle, R., Savage, C., Meyer, L. H., Sleeter, C. E., Hynds, A., & Penetito, W. (2011). Culturally responsive pedagogies in the visual and performing arts: Exemplars, missed opportunities and challenges. *Curriculum Matters*, 7, 26-47.
- Hodis, F. A., Meyer, L. H., McClure, J., Weir, K. A., & Walkey, F. (2011). A longitudinal investigation of motivation and secondary school achievement using Growth Mixture Modeling. *Journal of Educational Psychology*, 103(2), 312-323.
- Hynds, A., Sleeter, C., Hindle, R., Savage, C., Penetito, W., & Meyer, L. H. (2011). Te Kotahitanga: A case study of a repositioning approach to teacher professional development for culturally responsive pedagogies. *Asia-Pacific Journal of Teacher Education*, 39(4), 339-351.
- McClure, J., Meyer, L. H., Garisch, J., Fischer, R., Weir, K. F., & Walkey, F. (2011). Students' attributions for their best and worst marks: Do they predict achievement? *Contemporary Educational Psychology*, 36, 71-81.
- Meyer, L. H., McClure, J., Walkey, F., Weir, K., & McKenzie, L. (2009). Secondary student motivation orientations and standards-based achievement outcomes. *British Journal of Educational Psychology*, 79, 273-293.
- Meyer, L. H., Weir, K. F., McClure, J., Walkey, F., & McKenzie, L. (2009). *Longitudinal research on the NCEA and student motivation and achievement: A three-year follow-up.*

- Final Report to the Ministry of Education*. Wellington: Ministry of Education and Victoria University. Available at: <http://www.educationcounts.govt.nz/publications/schooling>
- Penetito, W., Hindle, R., Hynds, A., Savage, C., & Kus, L. (2011). The impact of culturally responsive pedagogies on students and families. In C. E. Sleeter (Ed.), *Professional development for culturally responsive and relationship-based pedagogy* (pp. 139-161). New York: Peter Lang.
- Sanders, P. (2012). Schools out – Gone to the beach: How some year 11 New Zealand high school students perceive study leave. In M. Hodis & S. Kaiser (Eds.), *Proceedings of the Symposium on Assessment and Learner Outcomes* (pp. 204-220). Wellington: Victoria University of Wellington.
- Savage, C., Hindle, R., Meyer, L. H., Hynds, A., Penetito, W., & Sleeter, C. E. (2011). Culturally responsive pedagogies in the classroom: Indigenous student experiences across the curriculum. *Asia-Pacific Journal of Teacher Education*, 39(3), 183-198.

¹ This paper is based on a keynote address presented at the Symposium on Assessment and Learner Outcomes, September 2011, Rutherford House, Wellington, New Zealand, featuring a summary of key findings from research supported in part by the Ministry of Education and the New Zealand Qualifications Authority from 2005-2011. While I accept full responsibility for interpretations of these data reported here, the research itself represents collaboration with John McClure, Frank Walkey, Kirsty Weir, Lynanne McKenzie, James Graham, Flaviu Hodis, Mimi Hodis and Michael Johnston. My sincere thanks to each of them.

Full List of Symposium Contributors

Contributors	Paper Title
Tahera Afrin	<i>External and internal moderation: Other side of the story</i>
Helen Anderson & Lisa Maurice-Takerei	<i>Assessment of literacy capability in adults</i>
Ann Balcolme, Liz Everiss, & Margaret Brennan	<i>Formative assessment in practice in tertiary education</i>
Stephen Bargh	<i>Technologies – marking</i>
Roseanna Bourke	<i>Online self-assessment in tertiary education: Exploring ways to use self-assessment to sustain learning</i>
Roseanna Bourke & Carolyn Tait	<i>How do students use self-assessment to learn in a tertiary institution?</i>
Therese Boustead	<i>Value-added modelling: A New Zealand monitoring success</i>
Gavin Brown & Zhenlin Wang	<i>A complex web of consequence, emotion, and agency: Hong Kong tertiary students' drawings of assessment</i>
Sue Cherrington	<i>Self- and team-assessment of early childhood teaching practices through the use of video-stimulated recall</i>
Jill Clark & Trish Baker	<i>Assessing group work in student industry projects: Is fairness achievable?</i>
Jim Clarke	<i>Does scenario based content delivery as an additional resource improve student performance in first year university Human Bioscience students?</i>
Graeme Cosslett & Rachel Dingle	<i>Smart tools to help learn about student and teacher motivation</i>
John Dickie & Mary Jane Shuker	<i>Popular culture, overall teacher judgments and national standards</i>
Frances Edwards	<i>Developing 'assessment capable' secondary teachers</i>
Angela Feekery & Lisa Emerson	<i>Assessing the learning process – promoting deeper source evaluation and justification of source use in first-year tertiary assessments</i>
Hilary Ferral & Charles Darr	<i>A step further with computerised adaptive testing – The 'snapshot' assessment</i>
Kay Fielden & T. Roper	<i>Online assessment</i>
Lester Flockton	<i>Directions for Assessment in New Zealand (DANZ)</i>
Dawn Garbett	<i>Assessing collaborative group work</i>
Dawn Garbett	<i>Assessing critical thinking in pre-service TE</i>
Gwen Gilmore	<i>Reducing exclusions and improving attainment – how can that be? Evidence from a United Kingdom secondary school</i>
James Graham & Luanna Meyer	<i>Individual good vs collective good – Māori and Pacific student and whānau perspectives on achievement, motivation and NCEA</i>
Bali Haque	<i>NCEA as systemic change in New Zealand secondary education</i>
Trudy Harris	<i>Building the connection between student appraisals and professional development: The use of an evaluative rubric for the interpretation of appraisals data</i>
Rawiri Hindle & Jenny Horsley	<i>The Leaf in the Wind</i>
Rosemary Hipkins	<i>Design experiments for assessment of "21st century" learning outcomes</i>

Contributors	Paper Title
Rosemary Hipkins & Edith Hodgen	<i>Laying learning foundations to lift success at NCEA level 3</i>
Rosemary Hipkins, Edith Hodgen, & Sally Robertson	<i>National Standards, moderation challenges and teacher learning</i>
Alice Ho	<i>Primary school using assessment to benefit student learning and improve teacher practice</i>
Flaviu Hodis & Mimi Hodis	<i>Overall (total) score of an instrument: Friend or foe?</i>
Pamela Hook	<i>What kind of assessment for learning is appropriate in the age of Google and Wikipedia? Facebook and You Tube? Smart phones and text messaging? Twitter and blogging?</i>
Pamela Hook	<i>"What happens next?" is what happens next. (Student led assessment)</i>
Michael Johnston & Vernon Mogol	<i>Differential gender-based performance in internally and externally assessed NCEA standards</i>
Chris Joyce	<i>The challenges of measuring "thinking" as a learning outcome in science</i>
Mei Kuin Lai, Stuart McNaughton, & Selena Hsiao	<i>Reading, writing, numeracy across Years 4-9</i>
Elliot Lawes	<i>TIMSS and national standards in Maths</i>
David Lillis	<i>A comparison of attainment in NCEA with that of the previous normative assessment system</i>
Vicky Mabin & Stephen Marshall	<i>"It's assessment, Jim, but not as we know it!"</i>
Jennifer Mackrell and Christine Pallin	<i>Secondary assessment and the NCEAs</i>
Michelle McCulley	<i>Improving student performance and satisfaction by the development of visual summaries</i>
Azra Moeed	<i>The impact of NCEA assessment on teaching, learning and motivation to learn science investigation</i>
Vernon Mogol & Michael Johnston	<i>An analysis of gender- and ethnicity-based differential item functioning in externally assessed standards for NCEA</i>
Blair Northcott	<i>Factors associated with developing conceptual understanding</i>
Patrice O'Brien & Mei Kuin Lai	<i>Year 10 reading comprehension. Year 10 student perspectives on achievement</i>
Alan Ovens & Nathan Brubaker	<i>Using individualised grading contracts in teacher education settings</i>
Elizabeth Peterson	<i>Socio-emotional key competencies</i>
Peter Sanders	<i>Schools out – Gone to the beach: How Year 11 students regard study leave for NCEA</i>
Michelle Schoenberger-Orgad	<i>Making it meaningful: Peer feedback to transform the experience of writing a literature review.</i>
Mark Sheehan	<i>The role of NCEA internally assessed research projects in teaching and learning history</i>
Boaz Shulruf	<i>Maximising the utility of the NCEA in university admissions</i>
Esther Smaill	<i>Moderating the national standards: teacher learning & assessment outcomes</i>
Louise Starkey	<i>Considering assessment in a digital cultural context</i>
Jim Strachan	<i>What do past advocacies and fit-for-purpose criteria say about evolving the NCEA qualification structure?</i>
Peter Tait	<i>Progressive use of data shown through ORBIT myLearning online profiling analysis</i>

Contributors	Paper Title
Kelvin Tan & Sarah Wong	<i>Coming to terms with formative and summative assessment</i>
Cherie Taylor-Patel	<i>Student-led conferences: What makes them work</i>
Mike Taylor & Andrea Milligan	<i>Conceptual understanding in NZ Geography education</i>
Pania Te Maro	<i>Mummy, why is Ranginui crying?</i>
Vicki Thorpe	<i>The assessment of group composing</i>
Martin Thrupp	<i>The RAINS project on national standards: Overview and initial findings</i>
Martin Turner	<i>Assessment and intrinsic motivation</i>
Brian Tweed	<i>Naturally occurring evidence and the new numeracy unit standards</i>
Gaye Tyler-Merrick & John Church	<i>The benefits of effective behaviour screening for children in schools</i>
Brenda Weal	<i>National Standards: Thinking wider about sources of evidence</i>
Lynne Whitney & Ngaire Bennie	<i>How does the Ministry of Education monitor its contribution to Māori achieving education success as Māori?</i>
Aaron Wilson, Stuart McNaughton, & Mei Kuin Lai	<i>Drop in achievement from primary to secondary</i>
Yang Yang, Eva Heinrich, & Elizabeth Kemp	<i>Assessment of online discussion: Collecting discussion content and generating analysis data for assessment and feedback.</i>
Colleen Young	<i>Progress thru NCEA student perspectives</i>
Tzu-Chieh (Wendy) Yu	<i>Medical student peer-evaluations during a general surgery objective structured clinical examination</i>