



Contemporary assessment practices in university: Impact on teachers and students

Professor Tony Harland, Dr Angela McLean,
Dr Rob Wass, Ellen Miller, Kwong Nui Sim

May 2014

Acknowledgments

This study was completed with the support of many people. We would like to express our gratitude to Ako Aotearoa for sponsorship and making the research possible, and to the University of Otago for genuine interest and encouragement in our work on assessment.

A very sincere and special thank you to all the academic staff and student participants who took part in the study and who volunteered their time, shared their experiences and helped us gain some understanding of the complex practice of assessment and grading. They truly made this research come to life.



This work is published under the [Creative Commons 3.0 New Zealand Attribution Non-commercial Share Alike Licence](https://creativecommons.org/licenses/by-nc-sa/3.0/nz/) (BY-NC-SA). Under this licence you are free to copy, distribute, display and perform the work as well as to remix, tweak, and build upon this work noncommercially, as long as you credit the author/s and license your new creations under the identical terms.

Table of Contents

Acknowledgments	i
List of Tables	iii
List of Figures	iii
Summary	1
Introduction.....	1
Aims	1
Key results.....	2
Implications.....	2
Introduction	3
Assessment and grading.....	3
Assessment at the University of Otago.....	4
History of structural change	4
The pilot study.....	6
Students' experiences.....	6
Lecturers' experiences.....	6
Study aim	7
Methods	8
The study	8
<i>Interview protocols</i>	8
Results	10
1 Students' experiences of assessment: overview.....	10
Theme 1 Number of internal assessments	10
Theme 2 Impact of grading on approaches to learning	11
Theme 3 Calculating risk as part of student identity	12
Theme 4 Independent study	14
Theme 5 Assessment fatigue	15
Theme 6 Formative evaluation	15
2 Teachers experiences of assessment.....	17
Theme 1 Awareness of assessment tasks	17
Theme 2 Assessment controls behaviour	17
Theme 3 The effect of the system	18
Theme 4 Small grades	19
Theme 5 Formative evaluation	19
Discussion	21
Assessment and learning	21
Integrated assessment	23
Work life balance.....	24
Conclusion.....	24
Implications for practice	25
A model to guide assessment practices.....	26
Limitations	27
References	28

List of Tables

Table 1. The purposes of assessment.....	3
Table 2. Study participants and disciplinary affiliations	8
Table 3. Estimated number of graded assessments per student each week	10
Table 4. Student's favoured assessments for learning.	12
Table 5. Student reading outside the syllabus	14

List of Figures

Figure 1: Model to guide assessment	26
---	----

Summary

Introduction

The primary aim of this study was to examine the impact of high-stakes summative assessment on student learning and teaching at the University of Otago. The term high-stakes summative assessment refers to marking student work when this counts towards the award of a degree.

The background to the research is the introduction of semester and modular organisational structures to New Zealand Universities throughout the 80s and 90s. At about the same time the practice of assessing students only by final exam largely disappeared and most courses in the tertiary sector adopted forms of frequent summative assessment throughout each semester, as well as final exams.

These radical changes to education have remained largely unexamined and the impact of frequent internal¹ assessment on student learning and teaching practice has received virtually no attention in higher education. Lack of inquiry into these matters is perplexing when some of the world's leading assessment researchers have, for some time, highlighted modular educational structures and frequent grading as a problematic issue for learning on the basis that these practices could be detrimental to some of the key educational outcomes required of a higher education (Rowntree, 1987, Rust, 2007).

The study builds on a pilot project conducted in 2011 (Harland, 2012). The pilot suggested that students were graded so frequently that this practice defined their whole educational experience and caused several challenges for teaching and learning. Teachers in the pilot found themselves in a position where they believed they had little choice in grading, even when practices were understood to be less than optimal. These results were of sufficient concern to warrant further investigation of assessment practices.

Aims

The present study used a large qualitative sample to ascertain if the pilot outcomes were reflected across the institution in different subject areas. The research question was:

¹ Internal assessment relates to coursework set during the semester as distinct from the end of semester examination.

If all teaching and learning experiences are shaped principally by summative high-stakes assessment, then what is the impact of this on how students and teachers experience higher education?

Key results

Academics rewarded all student work with grades, and students accepted this situation as normal. In some situations it was possible to experience hundreds of high-stakes summative assessments in an academic year. Students were fixated on grades and lecturers were not able to break the grading habit. Students would not work unless a percentage grade was attached to an assignment.

The problem with such an assessment regime is that it becomes much more difficult for the University to fulfil its educational objectives, to teach to its graduate profile and ensure all graduate attributes are attained. For example, how are we fostering autonomous learners if students will only apply themselves to graded assessments?

A majority of learning was fragmented in the sense that it was done in small segments that were often much less than a semester, paper or paper module. As such, quite insignificant quantities of knowledge were learned and never revisited. This 'building block' conception of curriculum (Ramsden, 1992) compartmentalises knowledge and provides an educational experience characterised by a surface approach to learning.

Both students and teachers were unsatisfied with the current assessment situation but neither group wanted radical change. Students were generally happy to accumulate small marks during a paper while being irritated and stressed by assessment arrangements. Teachers were aware of practices that would encourage deeper approaches to learning but felt trapped by a system and culture that did not allow them to move away from frequent high-stakes assessments.

Implications

Data indicated that changes could be made to benefit educational outcomes. The key recommendations for assessments that carry a mark or grade are:

1. Endeavour to decrease the number of assessments
2. Align assessment tasks with the most important outcomes for learning
3. Require larger more integrated assessments that test a range of skills, abilities and knowledge forms
4. Ensure that assessment leads to deep approaches to learning that also encourage autonomy, life-long learning, independence and responsibility as desired graduate outcomes
5. Do not use assessment solely as a behavioural mechanism that encourages students to work or motivates them to study

Introduction

Assessment and grading

This project is about looking at the impact of grading practices in an educational system in which all grades count towards the award of a final degree. This practice is called high-stakes summative assessment. It is worth starting with a reminder of why we assess and grade student work. Student grades are required so that the standards expected of a graduate at an institution can be set and so that students can be judged against each other. In general, grades are thought to be essential for the following purposes:

<i>Purpose</i>	<i>Explanation</i>
1. Measurement	Gathering information based on evaluative judgement
2. Formative evaluation	The use of information for learner improvement
3. Administrative decisions	Based on measurement and evaluation for making decisions that have a consequence

Table 1. The purposes of assessment (from the work of Astin and Antonio, 2013)

The key to good assessment practice is to understand the tension between purposes 1 & 2, and purpose 3. Measurement and formative evaluation ensure students are supported in their learning but there is always an impact on students from administrative decisions. These decisions have the potential to override the learning purposes of assessment and grading.

Examples of administrative decisions include passing or failing a paper, progression to the next level of study, access to restricted entry programmes, the final award of a degree, deciding on a grade point average (GPA), access to scholarships, and how students are portrayed with respect to potential employers or continuing postgraduate study.

Assessment at the University of Otago

Like other New Zealand universities, the University of Otago operates with a semester and modular structure that provides a liberal type of education for most students with much free choice of subjects and relatively few required papers for each designated degree. Each semester, students can take up to four papers and each is assessed. The content of papers is very rarely, if ever, re-assessed.

In addition, the University allows grades to be given for internal assessments as well as examinations. In this system every grade, from the day that a student enters the institution until their last exam before graduation, counts towards a GPA. In this sense, all assessment tasks are ‘high-stakes’.

What is of major interest in the present study is the idea of ‘over-assessment’, an expression encapsulating the notion that students are assessed and graded more than necessary. Although this term is in common usage, there are virtually no published studies that critique this phenomenon. Rust (2000) reported on the problems of modularity and frequent assessment for students at his institution, advising: ‘never do two modules with coursework in the same term’ (p. 126). In the only study that has looked at student experiences of frequent summative assessment, Trotter (2006) reports on students in a professional education course who were asked to do three pieces of linked assessment in one semester. In stark contrast, a pilot study at Otago suggested that students can be assessed and graded as frequently as every few days throughout their three years at university and such a regime must result in a particular type of educational experience of which we have little understanding.

Miller et al (1998) have speculated on the possible dangers of frequent assessment being allowed to dominate all other teaching activities and it has long been known that frequent assessment drives students to take a surface approach to their learning (Entwistle & Entwistle, 1991). Such a position would leave a deep approach to learning marginalised, and so is likely to put at risk certain higher order graduate attributes, such as critical thinking and the requirement to be a critic and conscience of society. Furthermore, one could speculate that, if students are continually complying with assessment demands to the exclusion of other curricula possibilities, there may be little space for effectively developing autonomy as self-directed or life-long learners.

These ideas, however, are untested. What has not been ascertained is the impact of frequent high-stakes grading practices on student and teacher experiences, and how these might vary in different disciplinary areas.

History of structural change

The traditional path to a first degree at the University of Otago was via year-long courses or ‘units’. By the 1980s, a number of departments had gradually introduced smaller teaching modules, which were known as papers or half-units. These smaller modules, however, led to complications for administering regulations and it was suggested that there should be changes to unify practices through the introduction of a paper system and a credit point system.

In 1992, most degrees and diplomas being awarded by the University were converted from a unit system to a paper/point system. The general guideline was that one point represented one hour of study per week for thirty weeks of formal instruction or independent study. The points assigned to each paper were specified in the appropriate schedule for each degree. Between 1994 and 1996 discussions took place about the introduction of a semester system that would support the new papers, and in 1997 the University adopted the present day semester system. Two factors relevant to the present research were noted at the time:

1. All semester papers were to be assessed by the end of the semester in which they were taught.
2. A full-scale review of assessment would be undertaken by Assistant Vice-Chancellors after considering a commissioned report by a working party on assessment and workload.

In the lead up to the changes, academic staff expressed concern that students might be over-assessed in the new system. In particular it was recognised that there should be a spread of assessments through each teaching period, and deliberation occurred surrounding the possibility of imposing limits on the number of summative assessment tasks in a paper with relative weights given to grades for internal assessments and final examination. Also discussed were the adequacy and timing of information given to students about assessment tasks.

In view of the concerns raised, Senate commissioned a review and a report was produced. The 'Review of Assessment Policies and Practices at Otago University' was based on an empirical study of assessment practices at the University (Gilmore, 1996). The report highlights many positive practices but notes that:

Excessive assessment was identified as a major issue which needs to be addressed by the University with some urgency (p. 1)

The report was influential in the drafting of the Senate Policy on Assessment of Student Performance that was released in 1997. Issues discussed with respect to this policy were the spread of assessments through each teaching period, limits on the number of assessment tasks, relative weights given to internal assessments and final examinations, the length of final examinations, and the adequacy and timing of information given to students about forthcoming assessment tasks. The statement on numbers of graded assessments has remained more or less the same since 1997:

8.1 The number of assessment tasks which individually count more than five percent towards a final grade should not exceed one third of the number of points associated with the paper. Thus, for example, an 18 point paper should not have more than six tasks counting over five percent in the final grade, including any final examination. In-course assessed tasks which are subject to plussage can be excluded from this limit, because students have the option of relying on their final examination performance.

There is no guideline for the number of internal (i.e. in-course) assessments worth less than five per cent. In other words, in the extreme, the policy allows up to 100 one per cent assessments per paper and, because the University's flexible choice modular system is embedded in semesters, it is near impossible to either calculate or manage an individual student's assessment load.

The pilot study

In 2011, a pilot study at the University sought the views of six 2nd year students and six of their teachers on how they experienced high-stakes summative assessment. This work was done in three sciences and one humanities department (Harland, 2012). It was observed that some students were assessed and graded so frequently, that all their study and learning was shaped in some way by this practice. The main findings from the pilot were:

Students' experiences:

1. Students were being frequently assessed and because of this they had no time to do any of the work required of them outside core graded-curriculum activities.
2. All students regularly missed teaching sessions in order to cope with assessment loads. They also co-operated in small groups to ensure that both assessment deadlines could be met and that course materials were shared.
3. Students who had high expectation for their grades felt they were always working at sub-optimal levels and felt they could do better work than that handed in. These students seemed more stressed than those with lower expectations, so the impact of frequent summative assessment may not be equal across the student body.
4. Students were stressed by the lack of co-ordination of assessment tasks between the courses they were taking (up to four a semester) and even between sections of the same course. For example, they found that several assessments could be due at the same time and that their lecturers were unaware of this.
5. Paradoxically students expressed a preference for having many small in-course graded assessments and felt that large assessments were too high-stakes. None wanted to revert to a final exam carrying 100% of the marks.

Lecturers' experiences:

1. Teachers did not know how many assessments students undertook as there was little communication between lecturers, departments and programmes.
2. Teachers were reluctant to reduce the number of assessments, despite experiencing high marking loads. The reason given was that students would then spend all their efforts on tasks in other papers that carried marks. This situation suggested assessment being used to control learning behaviour and

an 'arms race' between teachers, departments, subjects and modules. Grading controlled both student learning and student behaviour (Rowntree, 1987).

3. Teachers felt that they were under pressure from students to give marks for any submitted course work, even when they thought this might not be appropriate.
4. It was recognized that overall grades might not reflect overall performance when small marks were given for tasks.
5. Non-graded forms of assessment (i.e. formative assessment) were not used by academics.

The key message from staff was that since 1997 the University had entered into a new form of education without being fully aware of the impact of changes on students, teachers and learner outcomes. The important questions that arose from the pilot study were about the influences of frequent grading on student learning and educational experiences:

- What type of graduates are we producing from a system where all formal learning is done for a grade (Foucault, 1997)?
- Is the idea of 'reading for a degree' a thing of the past and as such we focus only on outcomes rather than study practices (Crook et al, 2006)?
- If we require certain graduate attributes such as autonomous learners, students who take responsibility for their own learning and students who take a deep approach to their learning, then how does frequent grading impact on these outcomes (see Barnett, 2011)?
- Does an assessment regime prevent students moving from needing assurance and approval from others to self-sufficiency (Hattie, 2009)?
- Do current assessment practices equip students for life beyond graduation (Boud & Falchikov, 2006)?

Study aim

The present research aims to:

1. Explore how students and academic staff experience a culture of frequent high-stakes summative assessment.
2. Inform practice and policy and directly stimulate organisational change.
3. Contribute to assessment theory and guide future research.

The guiding research question was:

If all teaching and learning experiences are shaped principally by summative high-stakes assessment, then what is the impact of this on how students and teachers experience higher education?

Methods

The study

Semi-structured interviews were carried out with students and academic staff at the University of Otago between February and July 2013. Sixteen academics and 46 students in their second or third year of study took part. Students were anonymous volunteers who responded to open requests for assistance in the research project. Teachers were all mid-career academics who were known to the research team and recruited by direct approach.

Discipline area of main interest	Academics	Students
Professional	5	15
Science	6	16
Humanities	5	15
Total	16	46

Table 2. Study participants and disciplinary affiliations

The University operates in divisions of Health Science, Science, Humanities and Commerce. Students can take papers across all four divisions and so they were categorised by their main interest and aspirations. Professional includes subjects such as Medicine and Law. Those from Commerce were included in Professional, Science or Humanities depending on their subject interests. Note that students in any three of the categories might be taking some papers in a different category.

Interview protocols

The 16 academics were given the results of the pilot study. The academic participants led the interview as they expressed their ideas around the issues raised in the pilot. The interviewer's role was to explore ideas raised and help the academic reflect on their thoughts.

For students, demographic data were collected with an indication of the final degree grade they aspired to. Six broad questions were then explored in a semi-structured interview:

1. Tell me about your experiences of being assessed at university?
2. How often are you assessed?
3. What do you see the advantages/disadvantages of internal assessment?
4. Do you ever do work that is not required for a grade?
 - a. Would you like to have the time to do other work?
 - b. Would you do this if you had the time?
5. Do you think your grades represent what you can do as a learner?
6. From what type of assessment do you learn most?

Interviews lasted about an hour and were transcribed verbatim. Student data were analysed using an inductive technique to extract key themes (Thomas, 2006). Data from academics are reported under the key outcomes of the pilot study. Data from students are reported under emergent themes.

Results

The results are reported in three sections. First, the students' experiences of assessment, second the academics' experiences, and third a comment on policy benchmarking. Each section begins with an overview, followed by presentation of the themes extracted from interview transcripts. Evidence is provided by way of illustrative quotes.

1 Students' experiences of assessment: overview

Data confirmed the pilot study in that students were fed up with being assessed so frequently but, when questioned about alternatives, preferred the status quo. When asked if they would like to revert to a single exam, virtually no one thought this was a good idea. Grading was linked very closely to motivation to study and students thought that without a grade attached to work, they simply would not do it. It was the key driver for them and they valued being driven, while being fearful that they may not be able to study without this type of assessment. So, they did not like the experience of being continually graded but they did not want to change.

Theme 1 Number of internal assessments

The data showed that many students were potentially faced with very large numbers of in-course graded assessments. These were often small in size and grade value but, as every single mark awarded at University counts towards a GPA, all were high-stakes. It was very difficult to ascertain precise numbers of graded assessments because the system does not allow for collection of this data. Table 3 shows the estimated number of assessments that students reported:

Course type	Number of students (2 nd and 3 rd year)	Avg. number of graded assessments/week	Range of graded assessments/week
Science	16	1.44	0.5 - 3.5
Humanities	15	0.98	0.5 - 3.0
Professional	15	0.68	0.0 - 2.0

Table 3. Estimated number of graded assessments per student each week

Students understood the tension between the number of assessments and their worth in relation to effort, learning and the administrative functions of assessment:

It seemed pointless doing ... the amount of work and time I put into these assignments and you'd get barely anything from them marks' wise which was good in that if you didn't do that well, it wouldn't be that big of an impact which is what I liked but at the same time, you felt like you were putting in so much effort and then you were getting barely anything in return, especially when you didn't do well. (Participant 13, p. 6)

In many subjects, integrating learning was difficult because teaching and curriculum were typically organized in segments (sometimes called modules) that are smaller than papers. Each segment is then graded independently.

I think having constant assessments with papers that aren't necessarily communicating with each other, there is a huge risk of overlap and a huge risk of overload for the students. It's very common that students have assessments almost due at the same time for very different disciplines and I think when students get overloaded, it tends to push them into a cram mentality so they just want to get the notes; they just want to get it done. They stop caring about quality. (Participant 47, p. 5)

Frequent internal assessment was seen by students as essential to motivate them to do the work and this was equated with learning, however numerous pieces of graded work in each paper caused problems with the timing of assessments, with many assignments being due in the same week, or on the same day. This was caused in part because course teachers did not or could not communicate with each other and plan assessment loads. This situation had negative impacts on students, including altering experiences of learning (Theme 2).

Theme 2 Impact of grading on approaches to learning

Being graded was a major concern for all students and they seemed to adopt a strategic approach to learning. Small assessment tasks were seen as trivial and stressful. However, they enjoyed collecting grades and were adept as grade counters and evaluating effort and risk in relation to reward.

I think if one was worth like five percent and one was worth 15 or 20, I'd start the 15 or 20 percent one a lot earlier than I'd start the five percent one because I feel like they'd expect a lot less, from the five percent one and they would mark it a lot easier because they can't expect the same level, like the same standard for each because why are you getting more marks for one and less for the other. (Participant 27, p. 3)

I'm not thinking about my learning... all I'm thinking is, what do they want to read? What do they want to hear? What do they expect from this? And what grade am I going to get and how's that going to affect my overall grade? (Participant 27, p. 11)

In addition, terms were not being used wisely to regulate student-learning behaviours. ‘Terms’ is a word used at the University of Otago and is equivalent to *course requirements* that allow a student to progress in their studies. Such requirements do not involve a grade that counts towards a degree. For example, a terms requirement might typically be to pass a formative test or to attend a number of classes and complete set work.

Students paid particular attention in their calculations when work carried higher marks and, importantly, it seemed these larger tasks changed their approaches to learning. They were asked what types of assessed tasks they learned most from:

Size of grade	Type of task students learned most from	Approach to learning required	Number of students
Larger	Inquiry task, research project, applied, work related project	Deep	39
Smaller	Short answer, frequent tests	Surface	6

Table 4. Type of assessment task described by students as the one they learned most from. These tasks have been grouped in terms of the size of the task and then associated with the theoretical required approach to learning. A deep approach is characterised by understanding and a surface approach by memorisation.

Theme 3 Calculating risk as part of student identity

Identity was often expressed in terms of a grade (“I am a B-grade student”) and not through subject or a passion for learning. Students carefully weighed and calculated the time required for a task and the reward of a grade and all would switch their efforts to the task that carried the most regard in terms of percentage points; the higher the reward, the greater the effort.

I mean if I got a C, you know, I don’t know what I’d do. I would seriously question my own ability as a person, as a student. (Participant 12, p. 26)

When you have a grade in the horizon that you need to achieve, you kind of tailor your study and efforts towards that more than just sort of I’m doing this or I’m doing that... when you’re aiming for a specific grade, that kind of, I think, almost dictates how much effort you put in. (Participant 47, p. 4-5)

If something’s only five percent and I’ve got a 15 percent test or whatever, I’ll just flag the assignment completely and study for the test. (Participant 17, p. 10)

Because students were reminded constantly that they had a high stakes assessment tasks hanging over them, there were rarely any spaces where they could put assessment out of their minds. There was no freedom to learn independently and surprisingly little genuine space for relaxation and social activities. Such concerns were particularly acute in professional programmes that had selective entry.

Even if I wasn't doing my assignment, it was still kind of like hanging over me. (Participant 13, p. 3)

Yeah, and just like always worried about something else. So like if I did have a day off, I'd feel bad for like doing the readings because I was, I'd be thinking, I should be doing my assessment... (Participant 20, p. 8)

My balance is far more university work and less 'me' time. It affects the quality of the time that I have for myself because I'm always thinking about university when I am doing those things. So I'm not fully there and then when I'm at university, I'm tired and jaded and not fully focused because I'm fatigued. (Participant 25, p. 10)

Many accounts of assessment experiences used emotive language around the fear of failure. Careful calculations were made during the semester that provided students with knowledge about what was required at that point to succeed. A few simply gave up if the desired final grade was no longer achievable.

The deadline is there, to me, as a fear. It's not a motivator. It's, it's fear. It's like I have to get this done or I'm going to fail. (Participant 12, p. 4)

There are those who do [assessment] because there is sort of competition... so once you know you've lost the game, people stopped trying (Participant 11, p. 10)

You know, if you have these three different things going, you can only do better at one if you do worse at another and that's the thing that's frustrating for me. (Participant 12, p. 11)

Nearly all students could be categorised as grade-obsessed and all could relate experiences of learning to percentage marks. They had preferences for the balance between internal assessment and exams (anywhere between 40 and 60% for the exam) and a strong sense of calculated fairness: they needed to know what they had to do for a grade and the reward had to match the task in terms of size and time required.

I'm definitely aware of [grades] and I would, I do kind of try and work out a bit like how much I need to get to certain grade, like overall for the paper. Umm, but probably for the first assessment, not too much because it's more after it, I'm like ohh okay, I got that grade. What do I need to get in the next one kind of thing? (Participant 54, p. 3-4)

Theme 4 Independent study

Almost without exception, students said that they did very little study or work outside of a graded assessment. In fact, we found that during the interviews, the concept of independent work was difficult for students to understand. When the idea of going beyond the syllabus was explained, they reflected thoughtfully and in general thought it might be something good to do if they had the time.

I think I'd like try and do lots of readings, like because in the lectures they give us references of things that, like would sort of broaden our knowledge. So yeah, I'd definitely get into the readings but I just don't have time.
(Participant 20, p. 8)

Academic Goal	Discipline	Number	How often do you go 'beyond the syllabus'?		
			Often	Sometimes	Never
Top Marks	Professional	8	0	0	8
	Science	6	1	0	5
	Humanities	7	1	1	5
B's	Professional	7	0	0	7
	Science	5	0	0	5
	Humanities	6	0	1	5
Passing	Professional	0	0	0	0
	Science	5	0	0	5
	Humanities	2	0	0	2
Total		46	2	2	42

Table 5. Students who read for a degree outside of the prescribed syllabus

When compared to their aspirations for marks, it can be seen in Table 5 that those aiming for top marks read outside of the syllabus more often than those aiming for B grades, and no one whose ambition it was just to pass, studied in this way.

Interviewer: So do you ever have time to do work that isn't for a grade?

Participant: Ahh, no (laughs). No. Yeah, I wouldn't, wouldn't see the point especially, well during semester time anyway. I just think, oh well, I'd be better off doing something that is going to get me through uni. (Participant 16, p. 9)

Many subjects required students to do independent prescribed reading but the response to this was again calculated. Reading would be done if a grade was at stake and a comment was made about extra reading: "I don't know how they ever expect anyone to get through it all. Insane" (Participant 30, p. 11).

The small number of students who spent time studying outside of required formal work recognised that this would probably be to their disadvantage in terms of achieving their desired grades and it was seen as a 'sacrifice' in the sense that they could not have the university experience they wanted because of excessive grading. For some, reading for a degree was a 'guilty pleasure'.

Theme 5 Assessment fatigue

All students expressed feelings of being stressed because of frequent grading. The final exam was still the most stressful event but the continual pressure to perform for small marks carried its own type of stress:

Or when you're doing three or four papers, there's an assessment at least once or twice a week and also you can't really, you're just trying to stay afloat and trying to keep up with the next assessment and not putting the effort and time into it that you should be to be able to get good marks and it takes away from studying for the next one or learning the content that you're being taught in the lectures because you're just trying to keep up with the assignment load. (Participant 25, p. 5)

The way in which students described their university experiences suggested that they worked hard and long hours. There was no evidence in this study of the stereotypical student spending many hours on leisure activities. Such commitment and pressure to work seemed to provide a very poor impression of study-life balance. This outcome was seen across the whole sample, regardless of aspirations for the final GPA. However, the stress of having to perform was also seen by some as positive as it motivated them to work.

Most students appeared worn down by the grading regime. In addition, students were after fairness and the idea of being fair tended to be equated with the value of marks awarded in relation to the size and how long an assessment task took (rather than, say, the intellectual challenge).

If there's something little, it's sort of like, I think of it... mmm, is it worth all my time and effort just for like a two or three percent mark? Then unfortunately, there's like several of those small things that add up. (Participant 37, p. 5)

Students did not want more time for internal assessments. Instead, their concern was about the timing of assessments and they wanted better communication between lecturers on their course and between courses. Overall, most students would have liked fewer but only slightly larger internal assessments.

Theme 6 Formative evaluation

A key study outcome was that formative feedback had become very rare, and where it did occur, students tended not to engage with it.

Interviewer: Do you get much feedback?

Participant: No. That's one thing that I would've probably liked more.
(Assess 55, line 478)

Every assessment counts towards your grade when it would be nice to have a chance to practice and just get a feel for the way your lecturer marks beforehand. I mean I know it's more marking for assessors and they don't like doing that but I suppose they are paid to help us through our course and to teach us ... (Participant 30, p. 3)

Despite many students' request for more feedback, when students did have the rare option to complete a purely formative exercise and receive comments from academic staff, most tended not to give the task their full effort.

I didn't put that much effort into my formative one at the start because, I mean I did it to like a but because I knew it wasn't being assessed, I wouldn't like go back and proofread it and double check I did everything right which showed but at the same time, I didn't do that badly so it would have been kind of, I would have been okay if it counted towards my grade but I know there was a couple of people in my class who just didn't do it because, ohh it's not worth anything and I don't have the time. (Participant 13, p. 11)

There was little evidence that students either received or took notice of formative feedback. They handed in work and moved on and because work was generally not assessed again, feedback was seen as unnecessary. In this sense all learning experiences and assessment were summative and focused on the grade, GPA, passing an assignment and moving on to the next new task. Compartmentalising study and frequent graded assessment made feedback a much more difficult proposition.

2 Teachers experiences of assessment

The data confirmed the outcomes of the pilot study and results are reported using the five themes academics were asked to consider before interview.

Theme 1 Awareness of assessment tasks

Pilot: Teachers did not know how many assessments students were subject to as there was little communication between lecturers, departments and programmes.

Academics did not know how many assessment tasks students were undertaking in other papers. Within their own various departments, seven academics were aware of the pattern of assessments, whereas six were not. In addition, one academic knew there was a lack of information within the department but stated that the issues were too difficult to sort out.

What our right hand does, the left hand doesn't know, in the university...so I think we need to have more communication or connections with each other (Participant 41, p. 22)

Theme 2 Assessment controls behaviour

Pilot: Teachers were reluctant to reduce the number of assessments, despite experiencing high marking loads. The reason given was that students would then spend all their efforts on tasks in other papers that carried marks. This situation suggested assessment being used to control learning behaviour and an 'arms race' between teachers, departments, subjects and modules. Grading was a technology of control for both student learning and their behaviour.

Academics understood grading as the key driver of student behaviour and saw themselves in a type of arms race with colleagues: if they did not grade work, then students would withdraw their effort and put it into another graded task. Those interviewed perceived that most students would only learn something if they were required to for a summative assessment: if there is no grade attached then students will not 'learn' it.

The fact that we have to do weekly tests to get them to do their readings for a tutorial is really absolutely ridiculous. (Participant 39, p. 11)

I've had a situation that was very humbling and upsetting and disturbing where I tried an integrative lecture with a student class which wasn't assessed...and three quarters of the class got up and walked out, one of them saying, I'm sorry, but we can't afford to stay here. We're off to do our assignment. That was sobering. (Participant 6, p. 3)

[Assessment] is a motivator...they probably find if they're being assessed then they will do the work, rather than definitely won't...it's a way for us to force them in to learning something. (Participant 42, p. 11)

Several staff acknowledged that assessment practices seemed to promote a level of strategic thinking amongst students, for better or for worse.

... in the past, I saw more an eagerness to learn. Now I see an eagerness to earn points and that in part is driven by the changes in the nature of the assessments. When I started here, the students were eager to do the assessments because they recognised that they had a pedagogical value. Now they are not eager to do the assessments and they'll do them, many of them, only to the level needed to get the points. So it's a difference in reward structure from education to points (Participant 9, p. 3)

I see that as a key part of a degree, that wider training of time management and game playing and strategic thinking about where do I put my effort at the moment? I do it every day in my job. We all do [and] I think being strategic about what it takes to pass should be a key skill in a student. (Participant 7, p. 5)

Theme 3 The effect of the system

Pilot: Teachers felt that they were under pressure from students to give marks for any effort made in course work, even when they thought this might not be appropriate.

Teachers indicated pressure from 'the system' (i.e. colleagues, the department, or wider university and external compliance to perform in research) affected their judgement and practices around assessment. Several commented on what they perceived to be the effect of modularisation of papers. It was suggested that knowledge had become compartmentalised and that students' ability to integrate different ideas was reduced as a consequence:

From a pedagogical perspective the ideal would be to devise an assessment such that it achieves the intended learning aims. We hear a lot about that with the teaching and learning plans and the various assessment goals and graduate objectives from the university, but then are not given the latitude and freedom needed to devise assessments that actually will generate that. (Participant 9, p. 5)

This modularisation of learning which I find very worrying because I think that I've detected, in the last five to eight years, just from experience, is that the students seem to have less ability to be able to integrate. So the knowledge is becoming compartmentalised. (Participant 6, p. 2)

The way that the examinations have tended to go is that we have this compartmentalisation. This module of work is what will be assessed, and therefore the students are not focusing necessarily on developing their broader contextual knowledge and understanding. (Participant 33, p. 5)

We let pragmatism get in the way of our ideals. (Participant 38, p. 18)

It was widely believed that if a lecturer did not assess and grade a piece of work, the work would not be done and, if they did not grade, students would put their efforts into another paper. This phenomenon can be likened to an assessment arms race and could be part of the explanation for the proliferation of high-stakes internal assessments.

Theme 4 Small grades

Pilot: It was recognised that overall grades might not reflect overall performance when small marks were given for tasks.

Academics' perceptions were mixed as to whether small, frequent assessment carrying a small percentage or grade, and mark accumulation, was a positive practice or not. Some academics could see that students appreciated the chance to build marks, and this was beneficial to their overall approach to learning. For others, however, small assessment tasks did not promote effective learning in students:

I think they roll their eyes and go 'oh gosh, not another assessment'...I think that students, the way a university degree is regulated, they expect assessments so they just sort of roll with it. (Participant 31, p. 14)

Furthermore, some academics expressed concern over the effect of small assessments on critical thinking and skill-building over time

If you're trying to think about how students approach kind of difficult critical problems, I don't really see how these little bits can help...I think the bigger picture assessment is the only way to go because I can assess in that larger context their thinking processes. (Participant 10, p. 1)

Theme 5 Formative evaluation

Pilot: Non-graded forms of assessment were not used by academics

Academics did not tend to use non-graded forms of assessment, and some perceived that students are not engaging in post-assessment feedback.

Despite formative tasks being an important opportunity to 'learn through assessment', it seemed that such tasks were not broadly embedded in curricula. There were, however, a few examples.

I think if we had more time, I would really like to sit with each student and say, 'this is where you went wrong' after each assessment, but we don't have that kind of time. (Participant 41, p. 21)

My only sort of misgivings about, or disappointment with that paper is that we don't really have a chance to read the report and then hand it back to them and let them improve on things. (Participant 32, p. 13)

It was three large assessments increasing in value...there was a few components to the assessment and it went over a number of weeks, so they would hand in a draft, peer review each other, get marked on those peer reviews and then revise their final drafts and submit the final version. (Participant 40, p. 2).

3 A note on disciplinary difference

Grading and assessment practices reported by teachers and students varied between and within disciplines, and it was not possible to identify patterns or differences that were subject related. In general the sciences had more frequent assessments but similar examples were also reported in, for example, commerce and health sciences.

Discussion

The research question was: 'If all teaching and learning experiences are shaped principally by summative high-stakes assessment, then what is the impact of this on how students and teachers experience higher education?' The main study conclusion was that all learning was shaped by high-stakes summative assessment and that this had negative impacts on both students' learning and lecturers' teaching.

There was a clear picture that new assessment and grading practices had evolved since the introduction of the Senate Policy on Assessment of Student Performance in 1997. These have largely been accepted by the academic community without a full understanding of the impact of change on students' education. We saw an assessment 'arms race' between academics and between their programmes and the predicted proliferation of graded internal assessments as a result of semesterisation had transpired (see Gilmore, 1996). It appeared that academics and students were trapped in practices that were likely to require whole institution change for these to be transformed.

The benchmarking study of assessment policies showed little policy development in this area worldwide. The research team presumed that universities who did not specify numbers of graded internal assessments in their policy a) did not have assessment proliferation as a problem, or b) were unaware of it, or c) were happy with current assessment practices. Although the study was unique to the University of Otago, informal enquiries during 2013 suggested that all New Zealand Universities have similar paper and grading systems set in a modular and semesterised structure. As such it is likely that many of the research outcomes from the study are of relevance across the New Zealand tertiary sector.

Assessment and learning

Students focused more on the products of assessment rather than the process of learning. This focus came from a culture of frequent assessments that 'count' in the students' minds and academic transcripts. Maximising grades was a priority and this was done strategically, even when the cost was missing out on lectures and tutorials, or trading one assessment against another. Students associated graded assessment with learning and the general curriculum pattern was that of micro-modules, some lasting only a few days, with small packets of knowledge that would be learned, assessed and then never revisited. Such a pattern was originally meant to apply to complete papers (i.e. semester-long) but, because papers were often broken down into smaller pieces and taught by several lecturers, each lecturer then assessed their contribution to the course, separately from other modules. At the same time, there was

the belief from most academics and students that any student effort had to be rewarded with a grade. It has been shown that students experience:

‘Intense pressure to perform well and obtain good grades that will open the door for a promising career. These pressures on the outcomes of learning can overshadow students’ desires to learn and improve.’

(Shim et al, 2012, p. 163)

Academics, however, had sophisticated views about the purposes and practices of assessment, but felt trapped by the system they found themselves in. Part of the explanation is the ‘arms race’ argument, exacerbated by an inability system-wide to discern how many internal assessments each student was asked to complete, across all their papers.

However, there seemed also to be a problem with the potential loss of some desired learning outcomes and attributes expected of a university graduate. The University of Otago Graduate Profile expects all those who graduate from the institution to be, for example, life-long learners and self-motivated with the ability to work independently (University of Otago). Foucault (1997) questions the type of graduates we produce when all learning is done for a grade; there must be some doubt that graduates are fully prepared for life after graduation when they have had little opportunity to practice self-directed learning, read a subject, or take responsibility for their own education (see Boud and Falchikov, 2006).

Some academics perceived that the current assessment regime did not prepare students fully for engagement at university or for professional life after university. However, there was also a view that frequent graded assessments gradually move students towards autonomy and challenge students to engage in learning on a regular basis (see Zepke and Leach, 2010). One respondent, however, suggested that experiencing frequent assessment was the perfect preparation for a neoliberal world in which students would continue to be assessed, judged, and held accountable throughout their lives. It has been argued that the continuous demand for up-skilling in the workforce will require individuals to re-skill themselves:

‘It is better, therefore to keep one’s learning superficial, as a deep understanding of a subject is likely to hinder rather than enhance career prospects.’

(Quill, 2011, p. 337)

Yet independence and autonomy seem to be pre-requisites for fully embracing a deep approach to learning (Barnett, 2011) and students were clear that they learned most from the larger types of assignments, such as essays and research projects. The binary idea that assessment drives a student to either take a deep approach or a surface approach (Rust, 2002) did not hold in the present study. Indeed, students adopted a strategic approach to their learning (i.e., do what is required) and we found that students would always adopt this, even when seeking a deep understanding. The strategy was to optimise grades at any cost.

‘Such behaviour, of course, may not be optimal (or ‘strategic’ in a broader sense) for the longer term, but may be a consequence of the instrumentalism implicit in political linkages between higher education and obtaining an appropriate graduate-level job, and in curricula where short-term achievements (in separate study units) are implicitly encouraged.’

(Yorke, 2006, p. 5)

Even though students enjoyed doing small tasks and collecting smaller grades, the vast majority favoured larger assessment tasks, and these are likely to encourage a deep approach to learning. Students were very good at calculating what they wanted to achieve or what was necessary, and so both deep and surface approaches were subservient to the dominant strategic approach. In other words meaningful learning, if necessary, would be sacrificed in order to achieve good grades. Academics expressed concern about students developing higher-order learning skills, such as critical thinking, and acquisition of these types of skills is likely to require more integrated forms of assessment.

Integrated assessment

Cutting down the number of in-course graded assessment tasks will change student experiences of learning. Most academics wanted smaller marking loads and the amount and type of assessment is something they had some control of in the University’s complex system. As a general rule, larger tasks that carry higher marks are typically more integrated and can more easily include feedback as formative assessment (even though there were examples in the study of subjects where small regular assessments were required, e.g., languages). If there were a shift to larger assessments, then students would still be strategic and change their approach while maximising grades. Then, if most academics moved in this direction it would change the character of the ‘arms race’.

A primary goal of integrated assessment is to encourage students to think of their educational experience in a more holistic way, instead of discrete compartments that do not always fit together (Ramsden, 1992). Current assessment practices at Otago require that a paper be fully learned and assessed within one semester and, when broken down into smaller modules, there can be a tendency to focus on propositional knowledge rather than procedural knowledge. As such students are asked to recall facts and theories and this typically represents a student’s mastery of material. While this factual-based knowledge is undoubtedly important to a programme and student learning, its assessment promotes the surface-approach to learning and tests more the ability to follow rules, memorise, and recall information verbatim. Such an experience may not provide an opportunity to see how this knowledge fits into the bigger picture of the discipline or allow the student to develop good study practices (Crook, et al 2006). In contrast, integrated assessment encourages a move away from mere memorisation of facts, towards a regime that requires students to demonstrate their comprehension of a subject by using knowledge of facts and concepts to seek out and discover links within and between topics, and to construct their own knowledge and understanding.

An integrated method would require carefully constructed pieces of high-stakes summative assessment so that assignments could not be answered by simply recalling facts. It would require students to analyze, synthesize, and draw conclusions based on their own educational and personal experiences (Ramsden, 1992). The currency of assessment would place value on the ability of students to discover depth in learning, to think critically, reason with knowledge, judge information, and synthesise concepts.

This type of integrated learning would need to be fully supported through knowledge preparation (combining lecture material with additional coursework activities, e.g. tutorials, practicals and independent reading). At the same time it should emphasize formative evaluation as feedback (Brookfield, 2006). According to Brookfield, feedback in this context has been more or less lost in modern assessment practices. Formative evaluation by lecturers and/or peers would require students to be engaged in a dialogue of diverse viewpoints and drive students to higher-order thinking without the fear of losing points or being ‘wrong’, emotions that were experienced strongly by students in the present study. By nature, the emphasis of making connections between knowledge forms, thereby viewing education as a whole in an integrated assessment model, would better develop student autonomy and foster higher-order learning outcomes.

Work life balance

All students complained about being “stressed” and we thought that it might be an overarching term used to represent a variety of emotional and physical responses, often with a negative connotation (see Robotham and Julian, 2006). Judging by their descriptions of how hard they worked, perhaps a more accurate representation would be feelings of ‘fatigue’ or ‘exhaustion’ (Law, 2007). Overall, there was little space for students to control their own academic lives because they were continually graded with too much assessment work to perform at their best. Stress in the context of learning can also be a useful motivator and drive efforts to a higher standard. However, we were unsure of this effect when a student was being continually stressed for three years.

Conclusion

In conclusion, this report has shown the impact of frequent high-stakes summative assessment practices on students and staff in several different ways through the analysis of key themes. Both teachers and students recognised that summative high-stakes assessment practices could be better, but felt constrained in achieving change to the system. An important opportunity does exist to shift to fewer and more integrated internal assessments. With a refined integrated assessment culture, the central aim for undergraduate education should be to achieve autonomy in learning through developing self-directed learning and critical thinking skills. Such skills would better meet the outcomes expected of a university graduate. Any changes in assessment to further these purposes would likely require academics to re-visit their desired learning outcomes, courses and possibly how they teach.

Implications for practice

The study describes a case of ‘over-assessment’. The practice change required to transform this situation is first to cut down on the number of internal high-stakes assessments. However, this strategy will not be relevant to all lecturers, subjects and students. For example, the research provided evidence that regular graded assessments in some subjects were necessary, for example, foreign languages. Nonetheless, any recommendations are intended as guidelines and general principles that are made in the context of seeking change across an institution. Without such an adjustment, then competition and inertia at the individual or paper level will ensure that current practices remain the norm. The recommendations below should be useful for any institution that recognises its internal assessment frequency and grading practices are not suitable for the desired outcomes for student learning.

It is important that there is clarity around learning outcomes and the attributes required of a graduate, and a good understanding of how graded assessments impact on these:

- 1 Maintain a balance between internal assessment and assessment through exams.
- 2 Reduce the number of internal assessments to a minimum and have larger tasks worth more marks.
- 3 Ensure learning is integrated in the task by requiring assessment of larger amounts and types of knowledge, skills and attributes typically expected of a graduate. For example, projects can be done in stages (with formative evaluation) but the final product receives the grade.
- 4 Focus assessment on the most important desired outcomes. For example, if critical thinking is seen as foundational to learning a subject, this should be assessed at each opportunity.
- 5 Motivate students to learn for reasons other than frequent assessment and fear of failure. A change will come about with opportunities to take a deep approach to learning.
- 6 Communicate assessment practices within and between papers. In a system of education that gives each student a great deal of free choice in what papers to study, coordination will be required between departments.
- 7 Ensure that students do not get all their assessments at the same time and that they have ample notice to help them plan.

8 Use course requirements (sometimes called terms) as a behavioural control rather than ‘busy work’ that carries a grade.

9 Create new spaces that will allow academics and students to provide feedback to help learning. Use self- and peer-review alongside lecturer comments, and offer opportunities to re-submit work before a final grade is given. In this way, assessment becomes a powerful learning opportunity.

A model to guide assessment practices

Based on the outcomes of this study, an example of a curriculum model to guide assessment practice was produced (Fig 1). Dislocated assessment (left-hand side of Figure 1) shows a paper in which there are four separate summative assessments and four exam questions based on different blocks of coursework and lecture material. Coursework and exams are compartmentalised because they are assessed independently. The integrated assessment option shows the same paper with an integrated staged internal assessment that draws on both coursework and lecture material, and focuses on the connections between lectures and topics. In a best practice situation, assignments would be handed in for formative evaluation and feedback, and then later summatively assessed. Knowledge and skills from internal assessment can be re-assessed as part of the exam that aims to test higher-order learning.

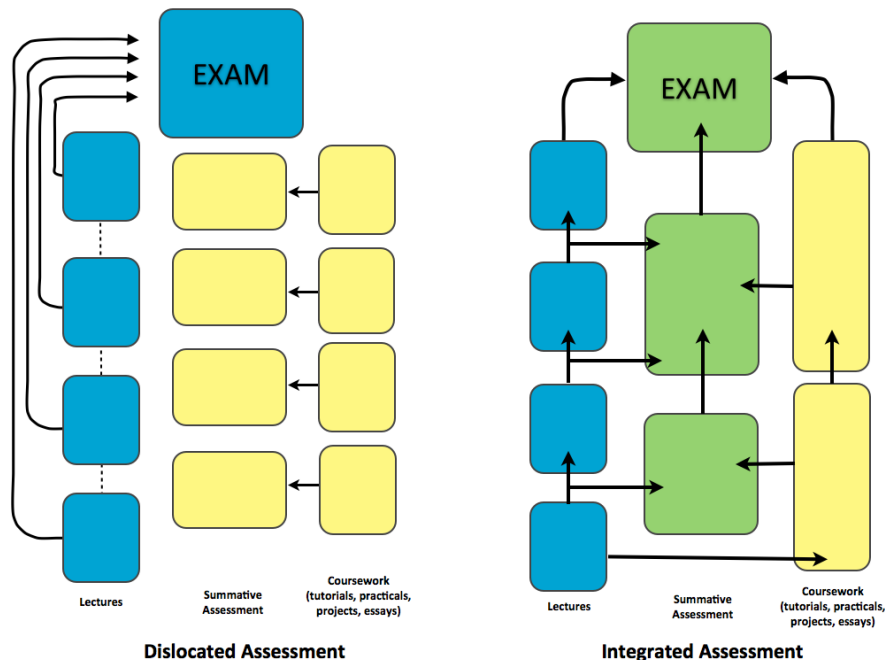


Figure 1: Model to guide assessment. On the left hand side the dislocated model shows four modules of coursework that are each graded. There are four lecture blocks and these are assessed at the end of semester in an examination, each with its own question. On the right hand side, the integrated model shows alignment between coursework and lectures, with work that is integrated with two assessments that are graded both internally and in examination.

Limitations

The main limitation of this study is that it is largely based on a qualitative account of teachers' and students' experiences and data cannot be connected strongly with learning outcomes. However we hope the evidence is convincing enough to encourage some careful thinking around assessment cause and effect, and its impact on teaching and learning.

A second limitation is that data interpretation was guided by the research team's idea about the desirable educational outcomes for a university education. We thought it important that a university should be aiming to:

- Encourage higher-order learning
- Help students become autonomous and self-directed learners
- Allow students to take intellectual risks without fear of losing marks
- Ensure they are motivated to learn for the love of learning
- Care for other's learning experiences in a community of learners

If one desired educational outcome is to produce calculating students who can strategically work out what is required to achieve a goal in a limited amount of time, (while responding to the fear of failure) then a regime of frequent internal high-stakes assessment works well. One academic respondent argued that such a system was good preparation for a neoliberal world of work centred on accountability, individualism and competition. We can understand this view but did not agree that this is the type of graduate we should be producing, or that our graduates should be trained to cope in a society as they find it. Instead we would expect our graduates to go out into the world to influence it, change it and make a difference. This was the perspective that underpinned how the data were interpreted and analysed and so it influenced how the study was conducted.

References

- Astin, A.W. and Antonio, A.L. (2013). *Assessment for excellence. The philosophy and practice of assessment and evaluation in higher education*, Maryland: Rowman and Littlefield Publishers Inc.
- Barnett, R. (2011). *Learning about learning: a conundrum and possible resolution*, London Review of Education, 9(1): 5-13.
- Boud, D. & Falchikov, N. (2006). Aligning assessment with long-term learning, *Assessment and Evaluation in Higher Education*, 31(4) 399-413.
- Brookfield, S.D. (2006). *The skillful teacher: on technique, trust, and responsiveness in the classroom*, San Francisco: Jossey-Bass.
- Crook, C., Gross, H., & Dymott, R. (2006). Assessment relationships in higher education: the tension of process and practice, *British Educational Research Journal*, 32(1): 95-114.
- Entwistle, N.J & Entwistle, A. (1991). Contrasting forms of understanding for degree examinations: the student experience and its implications, *Higher Education*, 22: 205-227.
- Foucault, M. (1997). *Society must be defended*, London: Penguin Books.
- Gilmore, A. (1996). *Review of assessment policies and practices at Otago University*, Dunedin: University of Otago.
- Harland, T. (2012). *Learning to teach in university: an introductory guide*, London, Routledge.
- Hattie, J. (2009). The black box of tertiary assessment: an impending revolution. In (Eds) L.H. Meyer, S. Davidson, H. Anderson, R. Fletcher, P.M. Johnston & M. Rees. *Tertiary Assessment and higher education student outcomes: policy, practice and research*, Wellington: Ako Aotearoa.
- Law, D.W. (2007). *Exhaustion in university students and the effect of coursework involvement*, Journal of American College Health, 55(4): 239-245.
- Miller, A.H., Imrie, B.W., & Cox, K. (1998). *Student assessment in higher education*, London, Routledge.

Quill, L. (2011). The disappearance of adulthood, *Studies in Philosophy of Education*, 30: 327-341.

Ramsden, P. 1992. *Learning to teach in higher education*. London: Routledge.

Robotham, D. & Julian, C. (2006). Stress and the higher education student: a critical review of the literature, *Journal of Further and Higher Education*, 30(2): 107-117.

Rowntree, D. (1987). *Assessing students: how shall we know them?* London: Kogan Page.

Rust, C. (2000). A possible student-centred assessment solution to some of the current problems of modular degree programmes, *Active Learning in Higher Education*, 1(2): 126-131.

Rust, C. (2002). The impact of assessment on student learning. How can the research literature practically help to inform the development of departmental assessment strategies and learner-centred assessment practices? *Active Learning in Higher Education*, 3(2): 145-158.

Rust, C. (2007). Towards a scholarship of assessment, *Assessment and Evaluation in Higher Education*, 32(2): 229-237.

Shim, S.S., Ryan, A.M. & Cassady, J. (2012). Changes in self-esteem across the first year in college: the role of achievement goals, *Educational Psychology*, 13(2): 149-167.

Thomas, D. (2006). A general inductive approach for analyzing qualitative evaluation data. *American Journal of Evaluation* 27, no 1: 1-10.

Trotter, E. (2006). Student perceptions of continuous summative assessment, *Assessment & evaluation in higher education*, 31(5): 505-521.

University of Otago, Interim Teaching and Learning Action Plan,
<http://www.otago.ac.nz/staff/otago027123.pdf> (accessed 5th November 2013)

Yorke, M. (2006). *Student engagement: deep, surface or strategic?* Keynote address delivered at the 9th Pacific Rim First Year in Higher Education Conference: Engaging Students, Griffith University, Gold Coast, Australia, 12-14 July.

Zepke, N. & Leach, L. (2010). Improving student engagement: ten proposals for action, *Active Learning in Higher Education*, 11(3): 176-177.