

Faculty and Students Conceptions of Assessment in Higher Education

Richard B. Fletcher · Luanna H. Meyer · Helen Anderson ·
Patricia Johnston · Malcolm Rees

Published online: 1 November 2011
© Springer Science+Business Media B.V. 2011

Abstract Assessment in higher education serves multiple purposes such as providing information about student learning, student progress, teaching quality, and program and institutional accountability. Yet, little is known about faculty and students' attitudes regarding different aspects of assessment that have wide-ranging implications for policy and practice in tertiary institutions. To investigate these views, parallel surveys of conceptions of assessment were administered to faculty and undergraduate students across four tertiary institutions including universities, an indigenous tertiary institution, and an institute of technology. A mean and covariance structures approach was used to test for measurement invariance and latent means differences between faculty and students regarding their conceptions of assessment. Results revealed differences in the latent means across the two groups. Faculty were likely to view assessment as a trustworthy process aiding teaching and learning, whereas students viewed assessment as focussed primarily on accountability and perceived assessment as irrelevant or even ignored in the teaching and learning process. These findings highlight the importance of ensuring that assessment policy and practices are fit for purposes, and are being carried out with integrity in ways that are transparent to and understood by both staff and students. While these results show how staff and students view assessment practices, one should keep in mind that while the

R. B. Fletcher (✉)
School of Psychology, Massey University Albany, Private Bag 102 904, North Shore Mail Centre,
Auckland, New Zealand
e-mail: R.B.Fletcher@massey.ac.nz

L. H. Meyer
Victoria University of Wellington, Wellington, New Zealand

H. Anderson
Manukau Institute of Technology, Auckland, New Zealand

P. Johnston
Te Whare Wānanga o Awanuiārangi, Whakatane, New Zealand

M. Rees
Department of Academic Vice Chancellor International, Massey University,
Private Bag 11-222, Palmerston North, New Zealand

sample was large and did incorporate different types of tertiary institutions, the inclusion of a broader range of disciplines would make the conclusions more generalizable.

Keywords Higher education assessment · Tertiary assessment · Latent means differences · Confirmatory factor analysis of conceptions of assessment

Introduction

Assessment in higher education performs multiple functions for varying purposes (Boud 1995; Carless et al. 2007). Assessments inform program selection decisions, determine student progression towards qualifications, and measure student learning towards attaining graduate profiles (Banta and Associates 2002; Brown et al. 1997). Assessments also provide information to faculty about teaching effectiveness (Biggs 2003; Ramsden 2003; Yorke 2003) and to students about how well they are doing and how they can improve their learning (Carless et al. 2007; Gibbs 2006; Gibbs and Simpson 2004). In addition, assessments are part of institutional quality assurance and accountability processes to validate the award of qualifications as well as the quality of disciplinary offerings, professional training, and the student experience (Knight 2002; Kuh et al. 2005; Maki 2004; Ramsden 2003). Boud (2000) discusses the challenges confronting assessment given its multiple and sometimes contradictory responsibilities: assessment is meant to inform student learning even as it sorts students into those who pass and those who fail; assessment measures learning outcomes but also compares students with one another; assessment should be objective and individually accountable but must evaluate the attainment of dispositions such as creativity, leadership and imagination (Knight 2006). These are complex and high-stakes responsibilities, yet James (2003) found that tertiary assessment was one of the least sophisticated aspects of teaching and learning. Consequently, Carless (2009a, b) recently called for work towards building trust in the integrity of assessment processes, while other higher education scholars highlight the need for developing a scholarship of assessment (Banta and Associates 2002; Price 2005; Rust 2007).

There is extensive literature over many years regarding the principles and best practice guidelines for higher education assessment to ensure that graduates master critical academic and professional outcomes (Banta and Associates 2002; Chickering and Gamson 1987; Hounsell et al. 1996; Maki 2004; Trigwell and Prosser 1991). What is less evident is the extent to which such principles and guidelines are empirically based on research regarding the different purposes and types of assessment, particularly as used, experienced, and viewed by those involved in the practice of assessment (Carless 2006). Attitudes towards and expertise in assessment by university faculty have an impact on the assessments they use, how assessments are incorporated into the teaching and learning process, and whether their assessment practices provide students with the opportunity to improve their performance (Swann and Ecclestone 1999; Tang and Chow 2007). Assessment attitudes and experiences by students will affect their approach to learning, whether they utilise assessment feedback in their future study, and the extent to which they develop the skills and understandings to become self-assessing lifelong learners (Boud and Falchikov 2006; Carless et al. 2006; Hattie 2009a; McDowell 1995; Tiwari and Tang 2003).

Recent research on conceptions of assessment—defined as one’s beliefs, meanings, and understandings of assessment—provides one basis for the investigation of the impact of assessment on learning (Brown 2004, 2006). Brown and Hirschfeld (2008), for example,

describe the impact of student conceptions of educational processes such as assessment on their educational experiences and learning. Similarly, Crooks (1988), Ramsden (1997) and Gibbs and Simpson (2004) noted how student learning is shaped in part by expectations modelled through assessment tasks, which in turn can be driven by faculty belief systems about the nature of learning and what is regarded as important in education (Biggs 2003). Teachers' views or theories of teaching regarding what must be learned influence their conceptions of assessment, what they choose to assess, and how they assess student learning outcomes (Watkins et al. 2005). Faculty have been shown to have mixed opinions on the purposes of assessment based on their attitudes about teaching and learning (McLellan 2004; Murray and MacDonald 1997). Those who viewed teaching and learning as the transmission of knowledge were likely to view assessment as a method to test students' ability to reproduce information. In contrast, those who saw teaching and learning as facilitating critical thinking viewed assessment as an integral part of the learning process for problem analysis and application (Samuelowicz and Bain 2002). McLellan (2004) has also identified that different understandings of particular assessment practices held by tertiary teachers are associated with strong coverage of disciplinary content and process but neglect some of the more analytical elements.

If student understandings of learning and student performance on assessments are influenced by faculty assessment practices, this could have a significant impact on whether or not intended educational outcomes are achieved. Hattie (2009b) points out that in higher education, "we implicitly trust our academics to know what they value in their subjects, to set examinations and assignments, to mark reliably and validly, and then to record these marks...and the students move on, upwards, and/or out" (p. 1). As students become more sophisticated in their understandings about newer forms of assessment and their own learning, they may not be satisfied with a model that relies on implicit trust. Boud's (2000) notion of "sustainable assessment" focuses on the imperative for approaches to equip students as lifelong learners rather than being passive recipients of assessment. Carless (2006) also emphasises the importance of engaging in dialogue with students about assessment processes so that there exist shared understandings about valued learning outcomes and mutual trust throughout the educational process. Without these shared understandings, there is the potential for distrust that can undermine not only assessment integrity but also the quality of the student learning experience (Carless 2009b).

Using Brown's (2004) research as a starting point, our aim was to investigate tertiary faculty and student conceptions of assessment in higher education. We were interested in how participants in tertiary education—both teachers and students—view assessment, and whether their views differ in ways that could have implications for how assessment is being and can be used in higher education.

Method

Participants

A voluntary sample of faculty teaching undergraduate programs, and students enrolled in the first year of a sub-degree or undergraduate program at four New Zealand tertiary institutions participated in the study. Across all participating institutions most major faculties were represented. The tertiary institutions included two universities, one polytechnic institute, and a wānanga (a tertiary institution by and for indigenous Māori, where either te Reo Māori or English may be languages of instruction). Each university enrolls over 15,000

undergraduate and postgraduate students across disciplines. The polytechnic and wānanga offer sub-degree, undergraduate and postgraduate degree programs, enrolling approximately 8,000 and 3,000 students, respectively. Following ethics review and approval, faculty and student were invited to participate in the study. The student sample comprised undergraduates new to the institution, and the faculty sample were those involved in teaching undergraduates. Participation was voluntary and confidential, and participants completed the survey during the second half of the academic year. A total of 877 faculty (males = 441: females = 436) and 1,224 students enrolled in a range of academic programs in their first year of their undergraduate degree (males = 379: females = 845) completed questionnaires.

Measure

Brown's (2006) abridged *Conceptions of Assessment* (CoA) measure was adapted from Brown (2004) for tertiary education contexts and administered to both faculty and students as part of a larger survey that also investigated other issues related to assessment. The CoA measure was designed for use in the compulsory school sector and its applicability to the tertiary sector is unknown. Thus, one of the aims of this study was to test the underlying factor structure within this education sector. The conceptions measured in the CoA are: assessment makes institutions accountable, assessment makes students accountable, assessment describes improvements in student abilities, assessment improves student learning, assessment improves teaching, assessment is valid, assessment is irrelevant and bad, assessment is irrelevant and ignored, and assessment is irrelevant and inaccurate. Respondents used a six point Likert scale ranging from strongly disagree to strongly agree. Previous studies of the CoA have shown it to have reasonable psychometric properties (Brown 2004, 2006). For example, a recent study conducted by Brown and Hirschfeld (2008) showed meaningful interactions between secondary school student conceptions of assessment and academic performance. Using a four factor measurement model—labelled student accountability, school accountability, assessments are fun, and students ignore assessments—they concluded that the robust measurement of students' conceptions of assessment was feasible. Brown (2004) also demonstrated empirically that primary school teachers held conceptions of assessment comprising four correlated major factors, which he labelled as conceptions of irrelevance, improvement, school accountability, and student accountability. Although Brown's (2004) research was focussed on the compulsory school sector, his paradigm of teachers' conceptions of assessment provides an empirically-derived starting point for examining conceptions of assessment held by higher education faculty and students. In studies conducted with New Zealand and Australian teachers, Brown (2006) found support for the hypothesised four-factor structure for the COA-III. The correlations, however, between some of the factors were high. For example, the factor correlations between the School Accountability and Student Accountability was $r = 0.58$, and School Accountability and the second-order factor Improvement was $r = 0.58$. Evidence for the underlying structure of the abridged CoA was similar to the long version. While some of the factor correlations are to be expected, one should consider carefully if the four factor model is the most appropriate for this measure.

Data analysis

A multi-group mean and covariance structures approach to invariance testing was used in these analyses and follows closely the methods outlined by Byrne (2006). First, a series of

confirmatory factor analyses (CFA) were calculated to determine the best fitting factorial model of the CoA. Second, the best fitting model was then calculated and tested for model data fit for male and female faculty, and for males and female students. Thirdly, factorial invariance and multi-group latent means analyses were conducted. All analyses were undertaken using EQS 6.00 (Bentler 2005).

CFA of the CoA

For both faculty and students, the CoA measure was tested with four competing models: (a) a correlated four factor model as specified by Brown (2006) with two first-order factors, these being institutional accountability and student accountability, and two second-order factors positive/improvement (improvement description, improvement students, improvement teaching, improvement validity), and negative or irrelevant conceptions of assessment (irrelevance bad, irrelevance ignored, and irrelevance inaccurate); (b) a three factor second-order model with the nine first order factors specifying conceptions of assessment for accountability (institutional accountability, student accountability), positive conceptions of assessment (improvement description, improvement students, improvement teaching, improvement validity), and negative conceptions of assessment (irrelevance bad, irrelevance ignored, and irrelevance inaccurate); (c) a two factor second-order model with the nine first order-factors specifying positive (institutional accountability and student accountability, improvement description, improvement students, improvement teaching, improvement validity) and negative conceptions of assessment (irrelevance bad, irrelevance ignored, and irrelevance inaccurate); and (d) a one factor second-order model with nine first-order factors (institutional accountability, student accountability, improvement description, improvement students, improvement teaching, improvement validity, irrelevance bad, irrelevance ignored, and irrelevance inaccurate). The best fitting model was then tested to establish baseline models across gender for both faculty and students before invariance tests were applied (Byrne 2006).

Invariance of the CoA using mean and covariance structures analyses

The aim of multi-group invariance testing is to determine if the factor structure of a measure is equivalent across different groups. Meredith's (1993) framework was used to test for measurement invariance by imposing a series of hierarchically structured tests ranging from weak to strong invariance. The process involves a sequence of increasingly stringent constraints imposed across the groups being compared (See Byrne 2006, pp 225–249 for a more detailed discussion).

Latent means analysis

If strong invariance is observed, then tests of multi-group latent means are appropriate. All latent means analyses retain the constraints from the preceding analyses. To test for latent means differences the designation of a reference group is required. This is an arbitrary choice as it is the sign of the difference test that is of importance; for all gender based analyses females served as the reference group. Any difference between the latent means is determined using significance tests based on a z -statistic and its associated probability value (See Byrne 2006, pp. 261–322 for a more detailed discussion).

Model fit indices

The Sartorra and Bentler (1988) scaled Chi-square statistic ($S-B\chi^2$) was used in favour of the uncorrected Chi-square as the data were multivariate non-normal. Mardia's (1970) normalised coefficient was 86.26 for faculty and 112.37 for students suggesting the data to be non-normal. Other fit indices employed were: the root mean square error of approximation (RMSEA; Steiger 1990) where values $\leq .05$ indicate good model fit and values $> .05$ and $< .08$ indicate acceptable model fit (Browne and Cudeck 1993); and the comparative fit index (CFI; Bentler 1990) where values $> .95$ indicate adequate fit (Mulaik et al. 1989). Because the χ^2 statistic has been shown to be sensitive to sample size and non-normality, we placed more emphasis on the change in CFI as suggested by Cheung and Rensvold (2002) to assess the fit of invariance models. Cheung and Rensvold (2002) suggested a change of $\Delta CFI \leq -.01$ between nested models for the CFI is indicative of the hypothesis for invariance and should not be rejected.

Results

Confirmatory factor analysis

The model fit of the CoA for both faculty and students suggested little difference between the two, three, and four factor models. Although there was some discrepancy between the various fit measures, with CFIs across both faculty and students being lower than accepted values, the RMSEA suggested an acceptable fit. We selected the two-factor higher order model as the best fitting for the further multi-group invariance tests based on statistical and theoretical grounds ($S-B\chi^2 = 1,290$, $df = 316$, $p = 0.05$, $CFI = 0.81$, $RMSEA = 0.06$). In part, our decision to not select the three and four factors model was based on the high correlation for the three factor model between accountability and improvement, which for staff was $r = 0.76$ and for students was $r = 0.91$ and for the four factor model the correlation for students between student accountability and improvement was $r = 0.76$. Generally across faculty and students the two-factor higher order model fitted similarly across gender (see Table 1).

The reliability estimates for faculty and students suggest low to high reliability, with values for the faculty ranging from $\alpha = 0.42$ to 0.84 and reliability estimates for students ranging from $\alpha = 0.48$ to 0.78 (See Table 2). Some of these estimates were on the low side; however, one should keep in mind that each first-order factor was comprised of only three items so that this is likely to be impacting some of these lower estimates and thus some caution is warranted when interpreting these scales. Of particular concern were two

Table 1 Within gender CFA for the two-factor second order model of the CoA for faculty and students

	df	S-B χ^2	$p \leq 0.00$	CFI	RMSEA
<i>Faculty</i>					
Females	315	868	0.01	0.80	0.064
Males	314	800	0.01	0.80	0.059
<i>Students</i>					
Females	316	1,198	0.01	0.83	0.058
Males	316	705	0.01	0.82	0.057

Table 2 CoA factors loading and reliability estimates for faculty and students using a sample of items

Factors and sample of CoA items	Faculty loadings	α	Student loadings	α
<i>Institutional accountability</i>	0.31	0.84	0.63	0.77
Assessment is an accurate indicator of an institution's quality	0.80		0.73	
Assessment is a good way to evaluate an institution	0.89		0.82	
Assessment provides information on how well institutes are doing	0.73		0.64	
<i>Student accountability</i>	0.53	0.40	0.76	0.47
Assessment is assigning a grade or level to student work	0.20		0.45	
Assessment determines if students meet qualification standards	0.94		0.82	
Assessment places students into ranks	0.08		0.18	
<i>Improvement describe</i>	0.98	0.61	0.99	0.68
Assessment establishes what students have learned	0.65		0.70	
Assessment measures students' higher order thinking skills	0.61		0.68	
Assessment is a way to determine how much students have learned from teaching	0.48		0.54	
<i>Improvement student</i>	0.89	0.74	0.95	0.76
Assessment feeds back to students their learning needs	0.66		0.71	
Assessment helps students improve their learning	0.80		0.78	
Assessment provides feedback to students about their performance	0.61		0.63	
<i>Improvement teaching</i>	0.76	0.48	0.94	0.59
Assessment is integrated with teaching practice	0.57		0.56	
Assessment information modifies ongoing teaching of students	0.57		0.65	
Assessment allows different students to get different instruction	0.44		0.49	
<i>Improvement validity</i>	0.69	0.80	0.77	0.73
Assessment results are trustworthy	0.77		0.74	
Assessment results are consistent	0.72		0.65	
Assessment results can be depended on	0.80		0.69	
<i>Irrelevant bad</i>	0.95	0.64	1.00	0.61
Assessment forces lecturers to teach in a way that is against their beliefs	0.59		0.51	
Assessment is unfair to students	0.65		0.65	
Assessment interferes with teaching	0.58		0.60	
<i>Irrelevant ignored</i>	0.86	0.56	0.75	0.54
Assessment results are filed and ignored	0.73		0.65	
Assessment has little impact on teaching	0.35		0.32	
Lecturers conduct assessments but make little use of the results	0.63		0.71	
<i>Irrelevant inaccurate</i>	0.79	0.48	0.84	0.57
Assessment results should be treated cautiously because of measurement error	0.61		0.52	
Assessment is an imprecise process	0.55		0.64	
Lecturers should take into account the error and imprecision in all assessment	0.33		0.41	

Italics = second-order factor loadings

items from the Student Accountability factor which show very low factor loadings. For the second-order factors, the reliability estimates were higher for positive conceptions of assessment (Staff $\alpha = 0.77$ and Students $\alpha = 0.86$) than for negative conceptions of assessment (Staff $\alpha = 0.61$ and Students $\alpha = 0.63$).

Invariance tests

The results for the staff and student means and standard deviations are shown in Table 3. The hierarchical models used to test for invariance showed very little change when more stringent constraints were added to each successive model. For faculty males and females comparison of each model showed that the Δ CFI was within the limits suggested by Cheung and Rensvold (2002) and therefore invariance was achieved. The same pattern was also observed for both male and female students and for staff and student comparison groups. Thus across all pairs of comparisons strong invariance was observed which allowed for latent means analysis between groups to be tested.

Latent means tests

The fit statistics for the test of latent group mean differences (see Table 4) suggest acceptable fit with CFIs (0.80–0.84) being lower than the values specified by Mulaik et al. (1989). The RMSEA (0.056–0.061), however, suggested an acceptable fit.

Latent mean differences between faculty and students were noted on the first order latent means with the exception of the assessment for description factor (see Table 5). Faculty tended to have higher mean scores on the improvement student, improvement teaching, improvement validity, and irrelevant inaccurate factors. Students, on the other hand, reported high mean scores on the institutional and student accountability factors, as well as on the factors labelled irrelevant bad and irrelevant ignored. For the higher order factors, there were statistically significant differences in favour of faculty for the positive conceptions of assessment, whereas students scored higher for the negative aspects associated with assessment.

Only one latent means difference was statistically significant between male and female faculty, and this was related to the factor labelled improvement teaching, with females reporting higher latent mean scores. For male and female students, differences on the latent means were noted for females for the institutional accountability, student accountability, improvement describe, student, and teaching factors. Females also reported high latent mean scores for the positive conceptions of assessment.

Discussion

Understanding the relationships between psychological variables assumes accurate and valid measurement of the underlying theoretical constructs. Establishing factorial validity of the CoA was an important aspect of this study as it provided the basis for understanding how relationships function and how groups may differ in those understandings. Furthermore the use of latent means analyses moves beyond the more traditional approaches (ANOVA and MANOVA) to understand if and how groups differ by first establishing measurement equivalence of the construct and then conducting tests based on such invariance tests. Thus, one can have more confidence that any differences noted between groups are related more to substantive issues rather than for measurement reasons

Table 3 Means and standard deviations for staff and students and for gender

	Staff		Staff female		Staff male		Stud		Stud female		Stud male	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Institutional accountability	8.52	3.57	8.52	3.31	8.54	3.80	10.65	3.29	10.81	3.29	10.31	3.26
Student accountability	12.39	2.85	12.23	2.90	12.54	2.81	13.49	2.63	13.59	2.58	13.29	2.70
Improvement describe	13.45	2.60	13.36	2.62	13.54	2.59	13.20	2.90	13.40	2.85	12.76	2.95
Improvement student	14.77	2.46	14.81	2.47	14.72	2.46	13.86	2.99	14.01	3.00	13.55	2.91
Improvement teaching	13.27	2.47	13.29	2.48	13.25	2.47	11.86	2.96	11.93	2.96	11.70	2.92
Improvement validity	11.83	2.96	11.43	2.90	12.23	2.97	11.00	3.10	11.07	3.06	10.86	3.13
Irrelevant bad	7.26	2.86	7.35	2.78	7.16	2.94	8.15	2.88	8.11	2.85	8.24	2.96
Irrelevant ignored	7.65	2.86	7.72	2.87	7.60	2.85	9.17	2.90	9.10	2.88	9.32	2.93
Irrelevant inaccurate	11.36	2.73	11.25	2.72	11.44	2.72	10.79	2.89	10.73	2.85	10.94	2.94
Positive conceptions of Assessment (second-order)	74.38	10.78	73.77	10.89	74.99	10.64	74.11	13.49	74.89	13.35	72.47	13.47
Negative conceptions of assessment (second-order)	26.28	6.55	26.36	6.63	26.19	6.48	28.20	6.87	28.03	6.84	28.61	6.86

Table 4 Fit indices for latent groups means tests

	df	S-B χ^2	CFI	RMSEA
Faculty				
<i>Males versus females</i>				
First-order latent means tests	674	1,734	0.80	0.061
Second-order latent means tests	682	1,765	0.80	0.061
Student				
<i>Males versus females</i>				
First-order latent means tests	677	1,960	0.83	0.056
Second-order latent means tests	684	1,971	0.83	0.056
<i>Faculty versus students</i>				
First-order latent means tests	677	3,578	0.84	0.060
Second-order latent means tests	682	4,368	0.83	0.061

(Cheung and Rensvold 2002). This said, one should keep in mind that the samples were derived from different types of tertiary institutions and that not all disciplines were represented. Thus, some caution is warranted when utilizing these results.

Some comment must be directed towards some of the low estimates of reliability. It is difficult to gauge how these estimates compare to Brown (2004, 2006) who did not report these. The measure used in this study was an abridged version of Brown's CoA consisting of 27 items (three per subscale), and thus for some scales the estimates of reliability with so few items were acceptable. Some scales had limited support, and we would recommend that better items be constructed for future efforts to measure these particular dimensions more reliably.

Generally, there was support for the underlying factor structure of the CoA as being a two factor higher-order model across the various groups. This is not in concordance with Brown (2004), who noted a four factor structure emerging from his research in the compulsory schooling sector. The factor correlations for the four factor model suggested that there was a high degree of association between the factors, so that from a statistical perspective it was difficult to argue for this model to be retained. While the fit indices for the two factor higher order model were on the low side, the RMSEA pointed to acceptable fit. We placed more emphasis on the RMSEA as it penalises overly complex models, and given the nine first-order factor and two second-order factors being estimated then the RMSEA suggested the model was worth retaining. Furthermore, the stringent tests of invariance suggested that the underlying factor structure was not invariant over the various groups. Achieving strong invariance provided a robust platform to conduct the latent means tests.

Latent means differences between faculty and students

Faculty and student conceptions of assessment revealed differences on all factors except for the assessment for improvement aspect. The differences noted in the results reflect the magnitude of perceptions regarding assessment and are not reflective of the degree of engagement. In many ways these results are an important set of findings that clearly show where there are points of difference between faculty and students with regard to

Table 5 Z-statistics (SE) for latent means groups differences tests

CoA factors	Faculty versus student (students reference) z-statistics	Male staff versus female staff (females reference) z-statistics	Males students versus female students (females reference) z-statistics
Institutional accountability	-14.07 (0.074)*	-0.07 (0.074)	-2.11 (0.056)*
Student accountability	-2.43 (0.012)*	0.80 (0.009)	-2.76 (0.019)*
Improvement describe	1.29 (0.029)	1.22 (0.042)	-3.51 (0.042)*
Improvement student	6.80 (0.029)*	0.04 (0.037)	-2.68 (0.044)*
Improvement teaching	16.38 (0.037)*	-1.36 (0.042)	-2.18 (0.047)*
Improvement validity	5.50 (0.042)*	4.08 (0.062)*	-0.87 (0.060)
Irrelevant bad	-8.08 (0.039)*	-1.50 (0.066)	0.47 (0.047)
Irrelevant ignored	-11.17 (0.048)*	-0.74 (0.070)	0.91 (0.067)
Irrelevant inaccurate	3.53 (0.037)*	0.53 (0.063)	1.10 (0.052)
Positive conceptions of assessment (second-order)	5.73 (0.048)*	1.29 (0.075)	-3.29 (0.066)*
Negative conceptions of assessment (second-order)	-8.65 (0.052)*	-0.99 (0.075)	0.90 (0.073)

* $p \leq .05$

assessment. It is not surprising that such differences exist as assessment is often a contentious issue. Knowing where points of difference exist allows for a reconciliation of such perceptions through better teaching and understandings of assessment practices at tertiary institutions.

Faculty reported viewing assessment as a method for improvement of student learning while also informing their teaching practices. The differences noted on the student improvement, teaching improvement and the improvement validity subscales suggest that faculty intend assessment to be focused towards understanding student learning. For example staff more favourably endorsed statements such as *Assessment provides feedback to students about their performance* and *Assessment allows different students to get different instruction*. Assessment was also viewed as being trustworthy and consistent whilst at the same time recognizing the potential for some inaccuracy to be present. For faculty, assessment practices can be a highly visible component of their teaching, and the grades they forward for students are scrutinized and moderated by examination committees and even external examiners before official grades are recorded. It may be, therefore, that faculty are more likely to be influenced in their conceptions of assessment by the knowledge that the grades they give will be examined for consistency and validity by their peers, within the institution, and sometimes also across disciplines and institutions where an external expert is called upon to moderate semester results across the program. Students do not see these stages in the process, hence they may be more likely to harbour concerns about assessment fairness, accuracy, and consistency.

In contrast to faculty, students tended to more favourably view assessment as having a primary focus on accountability both at the student and institutional level, which may

reflect the fact that assessment is “high stakes” for students in driving basic decisions about grades and even whether or not they will obtain a qualification. The differences noted on the institutional accountability and the student accountability subscales reflect the perception that assessment serves as an indicator of accountability, to students and to the institution. With the current educational climate of tertiary rankings and student completions, is not surprising that students tend to favourably endorse items such as *assessment provides information on how well institutes are doing*, or *assessment is an accurate indicator of an institution’s quality*. With the rising popularity of tertiary education league tables, be they national and international comparisons, it is easy to understand how students may perceive such tables being aligned to assessment practices. In reality, assessment practices are rarely related to such results.

Students also rated assessment as often irrelevant, and considered that assessment could be unfair and was often ignored. These results may also reflect students’ concerns about assessment as having serious, high stakes implications for their future within a context wherein they may not understand fully how assessment decisions are being made. Indeed, the endorsement of questions such as *assessment results filed and ignored* and *assessment is unfair to students* suggests that students are more sceptical about assessment than faculty. This said, students’ emphasize the role of assessment in improving learning suggesting they may be highly motivated to develop better understandings about how assessment works and how they can use assessment results to improve their own learning—if given the opportunity. It is important for faculty to stress the importance of assessment and its role in learning to students.

Implications for assessment policy and practices

Differences in beliefs, meanings, and understandings about assessment held by faculty and students raise important issues for higher education. An absence of clear institutional policy, an empirical base, and lack of transparency around assessment practices can contribute to these differences, particularly for students who are on the receiving end of assessment. Because we were unable to locate a validated measure of tertiary faculty and student attitudes about assessment, we made the decision to adapt a measure of conceptions developed for the compulsory sector. Perhaps the relative dearth of systematic research on higher education conceptions of assessment may be a consequence of historical confidence in ‘expert judgments’ by academic faculty regarded as leaders in their fields. Such judgments no longer go unchallenged in today’s universities, in a context where mass public education is associated with increased accountabilities for faculty and high stakes for students and graduates. Van den Berg (2002) discusses the disjunction between practice and regulatory or institutional intent, and Price (2005) attributes the reality of such inconsistencies across policy and practice in part to a weak development of shared concepts. We concur with Rust (2007) and Banta and Associates (2002) who emphasize the need for urgency in the development of a scholarship around issues in higher education assessment.

Other challenges include evidence that faculty-espoused values about assessment can be contradicted by actual practice (Eley 2006; MacLellan 2001). Meyer et al. (2010) reported complementary but also contradictory institutional policy and practice within a sample of tertiary institutions, including a lack of vision and clear sense of purpose among senior academic managers responsible for the integrity and validity of assessment decisions. Divergent views about assessment among faculty and students may be the unfortunate

consequence of the absence of clear principles for policy and practice in this area as well as an immature evidence-base. To address these issues, Meyer et al. (2010, pp. 348–349) present a practical framework for policy and practice that responds to three purposes of assessment (feedback on learning; selection and progression; and quality assurance/accountability) cross-referenced to key aspects of assessment approaches (integrity; equity; validity; manageability and utility). The diversity of conceptions of assessment is also likely to reflect diversity in prior experiences, so that more exposure to how assessment works for students and more transparency by staff about their assessment practices could mediate some of these differences.

There is a growing body of research underlying the importance of understanding views of assessment among faculty and students because of their impact on learning and the outcomes of learning (Bol and Strage 1996; Hattie and Timperley 2007; Knight 2002; Struyven et al. 2005). Student perceptions that assessment can be unfair and ignored in decision-making could reflect a lack of trust that has the potential to undermine student performance in educational programs and make it less likely that students will use assessment feedback to inform their future learning activities. In their comprehensive review, Struyven et al. (2005) found that attitudes about assessment have a strong impact on how students approach learning and what they learn. Further investigation is needed to demonstrate whether, and how, conceptions of assessment can be positively influenced and enhanced by institutional policy and practices that are informed by research (Price 2005; Rust 2007). Learners must develop the necessary understandings and strategies to evaluate their own performance if they are to become lifelong learners (Boud 2000; Carless 2009b; Hattie 2009b). They will not be able to do this if the assessment approaches and feedback provided by tertiary teachers are not constructed to provide information to improve learning, not delivered in a timely way, do not clarify expected standards of performance, and are seen as embedded within decision-making that is regarded as unfair and inconsistent. The tertiary teachers and undergraduate students in our study from several institutions of higher education had different ideas about how well assessment was working and what it was intended to do. Assessment is high stakes for everyone involved, hence requiring that both teachers and learners have better understandings and expectations about its purposes and processes.

Acknowledgments This research was supported in part by Contract Number TLRI-9233 managed by the New Zealand Council for Educational Research.

References

- Banta, T., & Associates. (2002). *Building a scholarship of assessment*. San Francisco: Jossey-Bass.
- Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, 107, 238–246.
- Bentler, P. M. (2005). *EQS 6.1: Structural equations program manual*. Encino, CA: Multivariate Software.
- Biggs, J. (2003). *Teaching for quality learning at university* (2nd ed.). Buckingham: Society for Research into Higher Education/Open University Press.
- Bol, L., & Strage, A. (1996). The contradiction between teachers' instructional goals and their assessment practices in high school biology courses. *Science Education*, 80(2), 145–163.
- Boud, D. (1995). Assessment and learning: Contradictory or complimentary? In P. Knight (Ed.), *Assessment for learning in higher education* (pp. 35–48). London: Kogan.
- 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.

- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Policy, Principles and Practice*, 11(3), 305–322.
- Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validations of an abridged version. *Psychological Reports*, 99, 161–170.
- Brown, G., Bull, J., & Pendlebury, M. (1997). *Assessing student learning in higher education*. London: Routledge.
- Brown, G. T. L., & Hirschfeld, G. J. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy, & Practice*, 15(1), 3–17.
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & S. Long (Eds.), *Testing structural equations* (pp. 136–162). Newbury Park, NJ: Sage.
- Byrne, B. M. (2006). *Structural equation modelling with EQS: Basic concepts, applications, and programming*. New Jersey: Lawrence Erlbaum Associates.
- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31(2), 219–233.
- Carless, D. (2009a). Trust, distrust and their impact on assessment reform. *Assessment and Evaluation in Higher Education*, 34(1), 79–89.
- Carless, D. (2009b). Learning-oriented assessment: Principles, practice and a project. In L. H. Meyer, S. Davidson, H. Anderson, R. Fletcher, P. M. Johnston, & M. Rees (Eds.), *Tertiary assessment and higher education student outcomes: Policy, practice and research* (pp. 79–90). Wellington, NZ: Ako Aotearoa & Victoria University of Wellington.
- Carless, D., Joughin, G., Liu, N. F., & Associates. (2006). *How assessment supports learning: Learning-oriented assessment in action*. Hong Kong: Hong Kong University Press.
- Carless, D., Joughin, G., & Mok, M. M. C. (2007). Learning-oriented assessment: Principles and practice. *Assessment & Evaluation in Higher Education*, 31, 395–398.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modelling*, 9, 233–255.
- Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *American Association for Higher Education Bulletin*, 39(7), 3–7.
- Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58, 438–481.
- Eley, M. (2006). Teachers' conceptions of teaching, and the making of specific decisions in planning to teach. *Higher Education*, 51, 191–214.
- Gibbs, G. (2006). How assessment frames student learning. In C. Bryan & K. Clegg (Eds.), *Innovative assessment in higher education* (pp. 23–36). London: Routledge.
- Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, 1, 3–31. Available online at: <http://www.glos.ac.uk/departments/clt/lathe1/index.cfm>.
- Hattie, J. A. C. (2009a). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. London: Routledge.
- Hattie, J. (2009b). The black box of tertiary assessment: An impending revolution. In L. H. Meyer, S. Davidson, H. Anderson, R. Fletcher, P. M. Johnston, & M. Rees (Eds.), *Tertiary assessment and higher education student outcomes: Policy, practice and research* (pp. 259–275). Wellington, NZ: Ako Aotearoa & Victoria University of Wellington.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112.
- Hounsell, D., McCulloch, M., & Scott, M. (1996). *The ASSHE inventory: Changing assessment practices in Scottish higher education*. <http://www.palatine.ac.uk/themes/AASHE/>. Accessed 16 March 2009.
- James, R. (2003). Academic standards and the assessment of student learning. *Tertiary Education and Management*, 9(3), 187–198.
- Knight, P. (2002). The Achilles' heel of quality: The assessment of student learning. *Quality in Higher Education*, 8, 107–115.
- Knight, P. (2006). Assessing complex achievements. In I. McNay (Ed.), *Beyond mass higher education: Building on experience* (pp. 96–104). Maidenhead: The Society for Research in Higher Education and Open University Press.
- Kuh, G. D., Kinzie, J., Schuh, J. H., Whitt, E. J., & Associates. (2005). *Student success in college: Creating conditions that matter*. San Francisco: Jossey-Bass.
- MacLellan, E. (2001). Assessment for learning: The differing perceptions of tutors and students. *Assessment and Evaluation in Higher Education*, 26(4), 307–318.
- Maki, P. (2004). *Assessing for learning: Building a sustainable commitment across the institution*. Sterling, VA: Stylus Publishing.

- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*, 519–530.
- McDowell, L. (1995). The impact of innovative assessment on student learning. *Innovations in Education*, *32*(4), 302–313.
- McLellan, E. (2004). Authenticity in assessment tasks: A heuristic exploration of academics' perceptions. *University of Strathclyde, Glasgow, UK Higher Education Research & Development*, *23*(1), 19–33.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543.
- Meyer, L. H., Davidson, S., McKenzie, L., Rees, M., Anderson, H., Fletcher, R., et al. (2010). An investigation of tertiary assessment policy and practice: Alignment and contradictions. *Higher Education Quarterly*, *64*, 331–350.
- Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stillwell, D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin*, *105*, 430–445.
- Murray, K., & MacDonald, R. (1997). The disjunction between lecturers' conceptions of teaching and their claimed educational practice. *Higher Education*, *33*, 331–349.
- Price, M. (2005). Assessment standards: The role of communities of practice and the scholarship of assessment. *Assessment and Evaluation in Higher Education*, *30*, 215–230.
- Ramsden, P. (1997). The context of learning in academic departments. In F. Marton, D. Hounsell, & N. Entwistle (Eds.), *The experience of learning* (pp. 198–216). Edinburgh: Scottish Academic Press.
- Ramsden, P. (2003). *Learning to teach in higher education* (2nd ed.). London: Routledge.
- Rust, C. (2007). Towards a scholarship of assessment. *Assessment and Evaluation in Higher Education*, *32*, 229–237.
- Samuelowicz, K., & Bain, J. (2002). Identifying academics' orientations to assessment practice. *Higher Education*, *43*, 173–201.
- Sartorra, A., & Bentler, P. M. (1988). *Scaling corrections for Chi-square statistics in covariance structure analysis*. American Statistical Association 1988 proceedings of the business and economic section (pp. 308–313). Alexandria VA: American Statistical Association.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioural Research*, *25*, 173–180.
- Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education. *Assessment & Evaluation in Higher Education*, *30*, 325–341.
- Swann, J., & Ecclestone, K. (1999). Improving lecturers' assessment practices in higher education: A problem-based approach. *Educational Action Research*, *7*(1), 63–84.
- Tang, S. Y. F., & Chow, A. W. K. (2007). Communicating feedback in teaching practice supervision in a learning-oriented field experience assessment framework. *Teaching and Teacher Education*, *23*(7), 1066–1085.
- Tiwari, A., & Tang, C. (2003). From process to outcome: The effect of portfolio assessment on student learning. *Nurse Education Today*, *23*(4), 269–277.
- Trigwell, K., & Prosser, M. (1991). Improving the quality of student learning: The influence of learning context and student approaches to learning outcomes. *Higher Education*, *22*, 251–266.
- van den Berg, B. (2002). Teachers' meanings regarding educational practice. *Review of Educational Research*, *72*, 577–624.
- Watkins, D., Dahlin, B., & Ekholm, M. (2005). Awareness of the backwash effect of assessment: A phenomenographic study of the views of Hong Kong and Swedish lecturers. *Instructional Science*, *33*: 283–309.
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and enhancement of pedagogic practice. *Higher Education*, *45*, 477–501.