

ResearchSpace@Auckland

Version

This is the Accepted Manuscript version. This version is defined in the NISO recommended practice RP-8-2008 <http://www.niso.org/publications/rp/>

Suggested Reference

Brown, G. T. L., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. F. (2009). Use of interactive-informal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning and Instruction*, 19(2), 97-111. doi: 10.1016/j.learninstruc.2008.02.003

Copyright

Items in ResearchSpace are protected by copyright, with all rights reserved, unless otherwise indicated. Previously published items are made available in accordance with the copyright policy of the publisher.

NOTICE: this is the author's version of a work that was accepted for publication in *Learning and Instruction*. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in *Learning and Instruction*, Vol. 19, 2009 DOI: 10.1016/j.learninstruc.2008.02.003

<http://www.elsevier.com/about/open-access/open-access-policies/article-posting-policy#accepted-author-manuscript>

<http://www.sherpa.ac.uk/romeo/issn/0959-4752/>

<https://researchspace.auckland.ac.nz/docs/uoa-docs/rights.htm>

Running head: Secondary students' conceptions of assessment

Use of interactive-informal assessment practices: New Zealand secondary students' conceptions of assessment

Gavin T. L. Brown ^{a,*}, S. Earl Irving ^b, Elizabeth R. Peterson ^c,
Gerrit H. F. Hirschfeld ^d

^a *Faculty of Education & ^c Department of Psychology University of Auckland, Private Bag 92019, Auckland, New Zealand*

^b *Cognition Consulting Ltd, Private Bag 92617, Auckland, New Zealand*

^d *University of Muenster, Psychological Institute II, 48149 Muenster, Germany*

Abstract

This study aimed to investigate how students' conceptions of assessment related to one another, how students' defined assessment, and how student conceptions of assessment related to their definitions of assessment. A nationally representative sample of New Zealand secondary students ($N = 705$) responded to a 45-item Conceptions of Assessment inventory and a list of 12 assessment practices. Well-fitting measurement models were found. The more students agreed that assessment was to help them improve the more they associated assessment with teacher-controlled practices. Further, the more students perceived assessment as irrelevant the more they defined it as interactive-informal practices. Unlike recent assessment reform advocacy, the students associated teacher-controlled assessment practices with assessment for learning, while more student-oriented practices were conceived as creating a positive social environment that was irrelevant to learning.

Keywords: Conceptions of assessment; Assessment practices; Secondary education; Structural equation modelling

* Corresponding author.

E-mail address: gt.brown@auckland.ac.nz (G. T. L. Brown)

1. Introduction

Assessment is a powerful and present force in students' lives. Assessment is used for a number of purposes including individual certification, improvement of teaching, and feedback on the quality of learning (Peterson & Irving, in press). Given the diverse role of assessment, understanding what assessment means to students is likely to be of great importance. Of real interest are students' beliefs and opinions about the purposes and nature of assessment. Individual behaviour and outcomes depend on, among other things, the beliefs and intentions people have towards phenomena, as well as their understanding of the societal behavioural and attitudinal norms around the same phenomena (Ajzen, 1991). Generally, the more positive people are about a specific behaviour the more they believe they are capable of doing it, and the more they believe their intended behaviour is socially acceptable the more likely it is they will act in accordance with their intention (Ajzen, 2002). Thus, examining students' attitudes towards competing assessment purposes may help us uncover important factors in what students do before, during, and after assessment.

1.1. Conceptions of assessment

To understand how the world appears to teachers and students, research has been conducted into their 'conceptions'. Conceptions have been defined as «general mental structures, encompassing beliefs, meanings, concepts, propositions, rules, mental images, preferences, and the like» (Thompson, 1992, p. 130). A major problem in investigating people's conceptions is that human actions may be so habitual that the reasons or plans for those actions or behaviour may be deeply buried (Ajzen, 2002; Huber & Mandl, 1984). If we want to understand people's conceptions we need to expose such thinking through the use of questionnaires and similar external stimuli.

At secondary school-level multiple and conflicting conceptions of assessment have been reported. These conceptions include: assessment improves learning, assessment makes students accountable, assessment is negative because it is unfair or bad, and assessment is liked because it is fun or beneficial (Brown & Hirschfeld, 2007; in press).

1.1.1. Assessment improves learning

Pajares and Graham (1998) surveyed 216 Grade 8 students in language and arts classes and found that they wanted honest, comprehensible, and constructive feedback on how to improve. The students wanted to improve and conceived that feedback in response to an assessment should do this. In a similar vein, American high school students reported that good teachers, in contrast to poor teachers, regularly tested and provided feedback to students about learning (Olsen & Moore, 1984). Peterson and Irving (in press) used multiple focus groups of New Zealand secondary students and reported that the students primarily saw assessment as a way of getting feedback to help improve their learning. However, there were certain caveats, viz., the feedback needed to be focused, honest, based on formal assessment, and be given by teachers.

1.1.2. Assessment makes students accountable

Another dominant theme is that assessment makes students accountable. The existence of the student accountability conception can be inferred from the responses

students make to assessments that have high-stakes consequences for individual students. For example, Brookhart and Bronowicz (2003, p. 240) concluded that, due to the consequences attached to classroom assessments, the high school students they studied “were ‘playing the summative game’, which includes protecting one’s reputation, self-worth, and self-efficacy as much as possible”. Furthermore, Zeidner (1987) reported that Israeli students preferred assessment methods that they perceived as maximising their grades especially when those in-course assessments contributed to grades. Such reactions to assessment can be taken to suggest that students view assessment as something that makes them accountable for learning.

1.1.3. Assessment is negative and irrelevant

Additionally, there is evidence that students differ in how negatively they perceive assessment. Most students become increasingly negative towards assessment as they work their way through the system. For example, the attitudes towards assessment of 54 Australian students in their first year of high school became increasingly negative for two reasons. First, there was much more assessment in secondary school and, second, the teachers’ assessment decisions appeared to be subjective to the students (Moni, van Kraayenoord, & Baker, 2002). In interviews with 31 minority students (i.e., urban African American and Latino) in their last year of American high school Walpole, McDonough, Bauer, Gibson, Kanyi, and Toliver (2005) reported that the students regarded the high-stakes university entrance tests as unfair obstacles (partly because of its one-shot nature) upon their life chances. Younger English children, on the other hand, blamed themselves for poor results rather than the tests (Reay & Wiliam, 1999). Perhaps partially in response to the high-stakes of accountability assessment, tasks on which high marks were difficult to obtain were regarded as irrelevant by Israeli secondary students (Zeidner, 1987). Peterson and Irving (in press) found that New Zealand secondary students judged assessment to be fair (and by implication not negative or irrelevant) only if it was completed alone, under controlled conditions, and without any assistance or second chances. Thus, it seems fair to conclude that students are quite sensitive to assessment which they perceive to be unfair, bad, or irrelevant to them.

1.1.4. Assessment is liked

There is evidence that students simply prefer whatever system of assessment they experience, regardless of the merits or deficiencies of that system (Blaikie, Schönau, & Steers, 2004). Students may not really be in a position to evaluate assessment methods that are not in their experience, nor will they necessarily know that they may need to change their responses to various assessment practices (Entwistle & Peterson, 2004). Perhaps, students simply normalise and adopt the values, beliefs, and preferences of their teachers (Zeidner, 1992).

However, research into students’ preferences for alternative assessments (e.g., portfolios, projects, self-assessment, peer-assessment, and other non-examination assessments) shows that the assessments that have been positively evaluated by students were more authentic and thus made learning more realistic or powerful (Bloxham & West, 2004; Gulikers, Kester, Kirschner, & Bastiaens, in press; Sambell, McDowell, & Brown, 1997; Segers & Dochy, 2001). However, authentic assessments are not perceived in the same way by all students; for example, Gulikers et al. (in press) found that freshmen and senior students differed in their ratings of authentic vocational assessment tasks, with seniors requiring less explicit support.

1.2. Research program into students' conceptions of assessment

The research reported in this paper is the fifth study in a series of studies into what secondary school students think about the nature and purposes of assessment and feedback. The first two studies (Studies 1 and 2) had large samples of students who did short versions of the questionnaire which limited the range of conceptions that could be investigated (Brown & Hirschfeld, 2007; in press). These two studies demonstrated that four conceptions of assessment (i.e., assessment makes students accountable, assessment makes schools accountable, assessment is fun, and assessment is ignored) were meaningfully related to student achievement in reading. The student accountability conception positively predicted reading achievement scores ($\beta = .42$), whereas, the conceptions that assessment was fun ($\beta = -.24$), assessment was ignored ($\beta = -.14$), and assessment makes schools accountable ($\beta = -.27$) all negatively predicted reading achievement scores. These results were seen as consistent with self-regulation theory (Zimmerman, 2001). Students who saw assessment as a constructive force for personal responsibility gained higher grades, whereas those who agreed with three non-self-regulating conceptions—i.e., (a) saw assessment as a means of making schools or teachers accountable for assessment results, (b) did not take assessment seriously, or (c) ignored assessment—received lower grades. Furthermore, the measurement model of conceptions of assessment was found to be invariant across gender, ethnicity, and age (Hirschfeld & Brown, under review).

Studies 3 and 4 (Brown, 2006; Brown & Hirschfeld, 2005) extended the previous studies by using longer questionnaires that elicited responses to a fuller range of conceptions of assessment, but were restricted to the smaller samples of students involved in the Conceptions of Assessment and Feedback (CAF) Project. The new items were in part designed in response to findings from within the CAF project (Peterson & Irving, in press). In Study 3 student conceptions were related to their academic performance and in Study 4 student conceptions were related to two clusters of assessment practices that the students identified with the term 'assessment' (assessment is a teacher-led formal process and assessment is informal and interactive). Study 3 showed that the student accountability conception was the strongest positive predictor of academic performance ($\beta = .24$), providing further support for the second study's finding. Study 4 showed that only one cluster of assessment practices was influenced by students' conceptions of assessment. Specifically, the more students believed that assessment was fun the more they defined assessment in terms of interactive and informal practices. Disturbingly, the conception that assessment benefited classroom relations or was personally enjoyable continued to be negatively related to academic performance.

Our research program into students' conceptions of assessment to date has been restricted either to short inventories that do not cover a wide range of conceptions, or to small samples from which robust estimates could not be obtained. The aim of this study was to simultaneously investigate in a large nationally representative sample (a) the strength and interrelations of a wide range of conceptions of assessment, (b) the practices students use to define assessment, and (c) how conceptions of assessment influence the way students define assessment.

1.3. Hypotheses

Earlier studies had identified multiple conceptions of assessment, and independent confirmation of those factors was sought in this study. However, since

the relationship of the assessment conceptions to each other could not be identified earlier due to small sample sizes (Studies 3 and 4), or to limited range of items presented (Studies 1 and 2), the inter-relationship of factors had to be established. Similarly, while the structure of the assessment definitions had been explored (Study 4), sample sizes had been too small to accept those results as robust. Thus, the hypotheses tested in this study are somewhat tentative.

In keeping with the findings of Studies 1-3 and consistent with the work of researchers such as Peterson and Irving (in press), Zeidner (1987), and Walpole et al. (2005) it was hypothesised that students would identify four inter-correlated second-order conceptions of assessment, namely assessment makes students accountable, assessment makes schools accountable, assessment is fun, and assessment is ignored (Hypothesis 1). In keeping with the formative-summative dichotomy (Torrance & Pryor, 1998), it was hypothesised that students would identify two major assessment practices, namely interactive-informal assessment and teacher controlled assessment (Hypothesis 2). Finally, in keeping with theories of reasoned behaviour (Ajzen, 1991) in which behaviours are predicted by intentions and beliefs, it was expected that students' conceptions of assessment would predict the types of assessment behaviours or practices students associated with the term assessment. Combining previous studies, we hypothesised that test-like assessments would be predicted by student accountability conceptions, while informal and interactive assessment practices would be predicted by enjoyment of assessment (Hypothesis 3).

2. Method

2.1. Participants

A stratified, random sample of 100 schools was approached to participate. The sample was stratified to ensure representation by four criteria (proportion of sample shown in square brackets): school social economic status (high [37%], middle [32%] and low [31%]); size of school/no. of pupils (Large > 1500 students [38%], Medium = 900-1500 students [23%], and Small < 900 students [39%]); region (Upper North Island [70%], Lower North Island [15%], and South Island [15%]); and school type (co-educational [68%], single sex male [13%], and single sex female [19%]). The final sample consisted of 705 students from 31 secondary schools and provided at least 100 students in all but one cell of the four stratification criteria. All participants were in Years 9 and 10 (age: $M=14.14$; $SD=.96$). These year groups were chosen as they are prior to the years in which the official National School qualifications system is administered.

2.2. Instruments

2.2.1. Conceptions of Assessment inventory

The Conceptions of Assessment-Version V (SCoA-V) inventory consisted of 45 statements related to various complementary and competing purposes of assessment (sample items are given in Table 1).

Insert Table 1 about here

Students were asked to rate the degree to which they agreed with each item using a six-point positively-packed rating scale. The scale had two negative responses,

1 (strongly disagree) and 2 (mostly disagree), as well as four positive responses, 3 (slightly agree), 4 (moderately agree), 5 (mostly agree), and 6 (strongly agree). Students were asked to tick the response box that corresponded to their belief about each item. This positively-packed response system has been found to be effective in eliciting variance in responses when participants are inclined to be positively oriented towards a construct (Brown, 2004; Klockars & Yamagishi, 1988; Lam & Klockars, 1982). Since people appear to agree with plural contradictory conceptions simultaneously (Fodor, 1998; Green, 1971; Laurence & Margolis, 1999), we have strong grounds for thinking that students would be inclined to agree with all conceptions and thus a positively packed rating scale was warranted.

2.2.2. *Assessment Practices list*

Students were presented with a list of 12 different assessment practices covering a range of traditional testing, teacher grading of homework, teacher-student interactions, alternative performance assessments, self- and peer-assessments all of which are considered commonplace in New Zealand schooling. The students were asked to "Select the kinds or types of assessment that come to mind when you think of the word assessment" on a binary scale. The options given were:

- (1) An examination that takes one to three hours.
- (2) I score or evaluate my own performance.
- (3) My class mates score or evaluate my performance.
- (4) The teacher asks me questions out loud in class.
- (5) The teacher grades or marks or scores the written work I hand in.
- (6) The teacher grades me on a written test that he or she made up.
- (7) The teacher grades me on a written test that was written by someone other than the teacher.
- (8) The teacher observes me in class and judges my learning.
- (9) The teacher scores a portfolio of work I have done over the course of a term or school year.
- (10) The teacher scores me on an in-class written essay.
- (11) The teacher scores my performance after a conference or meeting with me about my work.
- (12) The teacher uses a checklist to judge my in-class performance.

In this way, insights into the relationship of conceptions with practices that define assessment could be examined. It should be noted that students' definitions of assessment may not be the same as actual practices students experience; nevertheless, they provides a provisional insight into how beliefs structure students' experiences.

2.3. *Procedure*

The anonymous questionnaire comprising both the (SCoA-V) inventory and the list of the 12 assessment practices was administered during a single lesson by a teacher assigned by the school. Administration took approximately 15 minutes and occurred near the end of the third term in a four term year. Potential confounding teacher effects are mitigated by having so many schools and teachers participating, nonetheless experimental procedures would be needed to control confounds attributable to teacher, subject, or timing. Detailed administration guides were given to each teacher, including instructions for the teachers to read out to their class. This process helped reduce any biasing effects administrators may have had on the student responses. The students and their parents consented to the students' participation and they were told that their responses would be confidential and their decision to participate would in no way affect their relationship with their teachers or their school.

2.4. Analysis

A combination of exploratory and confirmatory procedures was used in two stages (Anderson & Gerbing, 1988). In Stage 1, the measurement models for the students' conceptions of assessment and the assessment practices were determined. This was done using exploratory factor analysis (EFA) for the SCoA inventory and multidimensional scaling (MDS) for the Assessment practices list. The EFA of the SCoA inventory was performed using maximum likelihood extraction and oblique rotation (Osborne & Costello, 2005). The MDS was computed using Euclidean distances on the 12 ordinal measures (0=not selected, 1=selected). Acceptable fit for MDS analyses is imputed when Kruskal's stress is $\leq .05$ and the number of dimensions is not greater than four times the number of variables (Hair, et al., 1998). Confirmation of exploratory models for both constructs was carried out in AMOS (Arbuckle, 2005). In Stage 2, we used regression analyses, confirmed with a structural equation model (SEM), to identify how conceptions of assessment predicted definitions of assessment.

EFA began with four factors and sought robust first-order factors within each factor. Poorly fitting items (e.g., poor conceptual fit, negative error variance, non-significant regression weights, or high modification indexes) were removed. Inter-correlations between factors that were statistically non-significant ($\alpha < .01$) were identified and removed. The finalisation of the results reported here depended on finding a model which was theoretically meaningful, consistent with the results of earlier studies, had good fit to the data, and which made the most parsimonious use of items available.

There is much debate about appropriate standards for imputing model fit especially for multi-factorial, hierarchical models with large sample sizes. The χ^2 statistic is overly sensitive to large sample sizes, and the goodness-of-fit indices—e.g., Tucker-Lewis Index (TLI), comparative fit index (CFI)—are overly sensitive to complex models, that is, those with more than three factors or with hierarchical structure (Cheung & Rensvold, 2002). In contrast, the standardized root mean square residual (SRMR), the root mean square error of approximation (RMSEA) and gamma hat have been shown to be resistant to the impact of large samples, complex models, and model mis-specification (Hu & Bentler, 1999; Vandenberg & Lance, 2000). RMSEA values less than .05 indicate very good fit, while values as high as .10 may still be of interest and of value in developing our understanding of the real world (Steiger, 2000). Gamma hat (recommended as most resistant to model mis-specification by Fan & Sivo, 2007) values $> .95$ indicate excellent fit and those $> .90$ are good. Thus, in this study fit was considered acceptable if simultaneously RMSEA and SRMR $< .08$ and gamma hat $> .90$, and good if RMSEA and SRMR $< .05$ and gamma hat $> .95$. Furthermore, the proportion of variance explained by a relationship can be inferred from the squared multiple correlation (SMC) of factors.

3. Results

3.1. Measurement models

3.1.1. Conceptions of assessment

The initial exploratory factor analyses found seven factors with eigen values > 1.0 , explaining some 48% of variance. Inspection of item loadings assisted in

dropping 12 items from further analysis: two items had nearly identical wording to other items in the same factor, four items had no loading $>.30$, three items loaded on factors to which they had no valid conceptual link, and two items were removed because they had weak loadings on factors which already had >5 items. The eight irrelevance items loaded on one factor but were separated into two factors in CFA for conceptual reasons. Likewise, the six accountability items loaded on a single factor but were separated in CFA into two factors for conceptual reasons. The six teacher improvement items were drawn from two factors which had an inter-correlation of $.40$ but were treated as one factor in CFA for conceptual reasons. Thus, eight factors loading onto four second-order factors were tested in CFA.

Students' conceptions of assessment were hierarchically arranged as shown in Figure 1. The model fit was acceptable, $\chi^2(481, N = 705) = 1551.57, p < .001$; TLI = $.88$; CFI = $.89$; RMSEA = $.056$, 90%CI = $.053$ -. 059 ; SRMR = 0.06 ; gamma hat = $.92$. The model was consistent with previous survey results confirming Hypothesis 1 about the structure of students' conceptions of assessment. Based on responses to 33 items, a total of 8 first-order latent factors were identified; moreover, there were four inter-correlated second-order factors, each one of which had loadings from two first-order factors (Figure 1). All second-order factor inter-correlations were statistically significant except for the correlation between positive affect/benefit and irrelevance which was removed from the model. Note that the irrelevance conception had an additional negative linear path to the Personal Enjoyment factor. The positive inter-correlations between the second-order factors accountability, positive affect/benefit, and improvement suggest some overlap between these conceptions setting these apart from the irrelevance conception.

 Insert Figure 1 about here

The Accountability conception consisted of "Assessment makes students accountable" (Cronbach's $\alpha = .71$) and "Assessment makes schools accountable" ($\alpha = .63$) factors. The Improvement conception consisted of "Assessment is used by the student to improve learning" ($\alpha = .88$) and "Assessment is used by the teacher to improve students' learning" (Cronbach's $\alpha = .80$) factors. The Irrelevance conception consisted of "Assessment is bad" (Cronbach's $\alpha = .77$) and "Assessment is ignored" (Cronbach's $\alpha = .67$) factors, with a weak inverse loading on the "Assessment is personally enjoyable" factor. The Affect/Benefit conception consisted of "Assessment is personally enjoyable" (Cronbach's $\alpha = .75$) and "Assessment benefits the class environment" (Cronbach's $\alpha = .87$) factors. The regression weights from each factor to its second-order conception were robust, as likewise, were the paths from each item to its corresponding factor.

The positive inter-correlations among the Improvement, Accountability, and Affect/Benefit conceptions indicated that there was a tendency for students to respond to each conception in the same way. While improvement and accountability may be seen as antagonistic purposes, certainly for the students the correlations suggest that these three second-order conceptions have between half and one-third of the variance between them in common ($r^2 = .59, .41, \text{ and } .36$, respectively). In contrast, the correlations between the Irrelevance conception and the three other conceptions had much greater variability. There was a statistically non significant relationship with Affect/Benefit (and so removed from the model), a weak inverse relationship with Accountability, and moderately inverse relationship with Improvement. In other words, the irrelevance of assessment is most associated with it not providing

improvement-oriented information to either the student or the teacher. These data suggest there may be two dimensions within the inventory; a matter currently being investigated (Weekers, Veldkamp, & Brown, 2007 submitted).

To further understand this model, mean scores for each first-order factor were examined (Figure 2). The factors within each conception had very similar mean scores, with more than moderate agreement given on the mean to the two Improvement factors, slightly less than moderate agreement given to the two Accountability factors, about slight agreement given to the two Affect/Benefit factors, and disagreement given to the two Irrelevance factors. Taking the correlations and the mean scores together, it would appear that secondary students conceived of assessment as first being about improvement, second about accountability, while not being irrelevant and only a little being a positive social or personal experience. In other words, students appeared to be thinking: "Yes, assessment helps improve my learning and it holds me accountable, but it doesn't feel good. However, it is a bad thing when it doesn't help me, or my teacher, improve my learning."

 Insert Figure 2 about here

While not the focus of this study, it is possible to examine whether SCoA scores differed by demographic characteristics. Multi-factorial MANOVA examined the impact of school social economic status, size of school/no. of pupils, region, and school type simultaneously on the eight SCoA factors. There were no significant effects for size (Wilk's $\lambda=.97$; $F(16)=1.26$; $p=.21$), region (Wilk's $\lambda=.98$; $F(16)=.84$; $p=.65$), and school type (Wilk's $\lambda=.98$; $F(16)=1.10$; $p=.35$). While school social economic status was statistically significant as a main effect (Wilk's $\lambda=.94$; $F(16)=2.46$; $p<.001$), this applied to only two of the eight conceptions (i.e., Assessment is Bad and Student Accountability) and the adjusted R^2 for these two factors was only .04 and .03 respectively. CFA invariance studies are required to further investigate the nature of any differences in SCoA score attribute to school social economic studies.

3.1.2. Assessment definitions

The MDS analysis found three clusters in a two-dimensional solution (Kruskal stress =.05), one of which consisted of just the single Examination item. The other two large groups of practices were as expected the formal test-like and interactive-informal processes. The Examination item was much closer to the test-like cluster than the interactive-informal cluster. The three cluster solution was inadmissible in CFA, so the Examination item was joined to the test-like cluster. CFA analysis of a correlated two cluster model resulted in a well-fitting solution, $\chi^2(53, N = 705) = 103.72$; $p < .001$, TLI = .92, CFI = .93, RMSEA = .037, 90%CI = .026-.047; SRMR = .04; gamma hat = .99 (see Figure 3). While the internal estimates of cluster consistency were moderate at best (Teacher-Controlled Cronbach $\alpha = .51$; Interactive-Informal Cronbach $\alpha = .65$), the CFA procedure, by taking into account correlations, covariances, and residuals generates a more comprehensive indicator of scale quality. Thus, Hypothesis 2 regarding two major definitions of assessment practices was confirmed.

 Insert Figure 3 about here

The mean for each of the two assessment definitions ranged between 0 and 1, with 1 indicating that all of the assessment practices contributing to the factor were selected, while 0 indicated none had been selected. As shown in Figure 4, the mean for Teacher-Controlled assessment definition was .65 ($SD = .27$), while the mean for Interactive-Informal assessment definition was .24 ($SD = .24$). The correlation between the two assessment definitions was $r = .25$, clearly supporting the measurement model of two independent assessment definitions. Thus, we can conclude that students associated the term “assessment” predominantly with test-like practices such as examinations, essays, and teacher grading of tests and homework.

 Insert Figure 4 about here

While also not the focus of this study, it is possible to examine whether assessment definitions scores differed by demographic characteristics. Multi-factorial MANOVA examined the impact of school social economic status, size of school/no. of pupils, region, and school type simultaneously on the two assessment practices clusters. There were no significant effects for size (Wilk's $\lambda = .99$; $F(4) = 2.35$; $p = .05$), region (Wilk's $\lambda = .99$; $F(4) = .97$; $p = .43$), school type (Wilk's $\lambda = .99$; $F(4) = 1.44$; $p = .22$), and school social economic status (Wilk's $\lambda = .99$; $F(4) = .23$; $p = .92$). While this approach confirms stability of assessment practices definitions across school characteristics, CFA invariance studies are required to confirm this result.

3.2. Structural model

While it would be possible to analyse inter-correlations between the two instruments, it is more in keeping with the theory of reasoned behaviour to conduct regression analyses. The theory claims that beliefs about intentions predict behaviour and so Hypothesis 3 suggests that beliefs about the nature and purpose of assessment predict the type of practices students would define as assessment. Thus, to explore Hypothesis 3, we analysed how students' conceptions of assessment influenced the practices they defined as “assessment”. Multiple regression analysis using the assessment definition as univariate dependent variable was conducted with the eight factors for each assessment definition as predictors. Only four of the factors had a statistically significant or near significant loading on any assessment definition.

Interactive-Informal assessment practices were predicted by the factors Assessment is used by the teacher to improve students' learning (Teacher Improves Student Learning), $\beta = .12$, $t = 2.51$, $p = .01$, and Assessment is ignored (Ignore), $\beta = .10$, $t = 1.89$, $p = .06$. While the latter result would normally be considered non-significant, it is worth noting that SEM is better able to test for statistical significance of regressions because it examines all relations simultaneously. This model had an R^2 of .02, $F(8) = 2.15$, $p = .03$, and effect size of $f^2 = .02$.

The Teacher-Controlled assessment practices were predicted by the factor Teacher Improves Student Learning, $\beta = .14$, $t = -2.29$, $p = .02$, and inversely by the Assessment is personally enjoyable (Personal Enjoyment), $\beta = -.14$, $t = -3.02$, $p = .003$. This model had an R^2 of .04, $F(8) = 3.86$, $p < .00$, and effect size of $f^2 = .04$. Note that f^2 effect sizes between .02 and .14 are considered small (Cohen, 1992). Thus, there was a practically meaningful relationship between conceptions of assessment and the assessment definitions.

This result was confirmed in an SEM having four paths from the four identified factors to the two assessment definitions. The structural model (Figure 5) had good fit characteristics, $\chi^2(926, N = 705) = 2137.61$; $p < .001$; TLI = .88; CFI = .89; RMSEA = .043; 90%CI = .041–.046; SRMR = .06; gamma hat = .93. The Teacher-Controlled assessment definition was predicted positively by the factor Teachers use assessment to improve learning (Teacher Improves Learning) ($\beta = .32$) and inversely by the factor Assessment is personally enjoyable (Personal Enjoyment), $\beta = -.33$. The Interactive-Informal assessment definition was predicted positively by the factor Assessment benefits the class environment (Class Environment), $\beta = .14$, and by the factor assessment is ignored (Ignore), $\beta = .17$. In this model, the SMC as a measure of proportion of variance explained for teacher-controlled practices is 89%, while only 28% for the interactive-informal practices. Clearly, other unidentified constructs play a significant role in students' definition of assessment as interactive-informal practices; whereas the teacher-controlled practices are well-explained by these two conceptions of assessment

 Insert Figure 5 about here

This model revealed that as far as New Zealand secondary students were concerned, increases in the belief that teachers use assessment to improve learning and reductions in the assessment is personally enjoyable factors predicted an increased definition of assessment as Teacher-Controlled assessment practices. At the same time, an increase in the factors assessment benefits the class environment and assessment is ignored led to stronger definitions of assessment as interactive-informal practices. In other words, if students believed that assessment helped the classroom environment, and if they ignored assessment, then they were likely to define assessment as informal and interactive practices. Whereas, if students believed assessment was unpleasant (not personally enjoyable) and that it was used by teachers to improve their teaching, they tended to define assessment as a teacher-controlled practice.

4. Discussion

This study examined three hypotheses about students' conceptions of assessment, their definition of assessment, and the relationship of conceptions to definitions. Hypothesis 1 found support in an acceptably-fitting measurement model. The model was multi-dimensional, hierarchical, and had four inter-correlated conceptions, to which the students had varying degrees of agreement. The students mostly agreed with assessment improving learning, moderately agreed that it made schools and students accountable, slightly agreed that it had positive affect/benefit, and disagreed that it was irrelevant. They considered assessment that did not help students or teachers improve student learning to be irrelevant and otherwise positively associated improvement with accountability and social benefit purposes.

Hypothesis 2 regarding the definition of assessment as consisting of two factors was confirmed in a two-factor model. These were teacher-controlled assessments and interactive-informal assessments, with the students selecting considerably more of the former practices.

Hypothesis 3 regarding the relations between conceptions of assessment and assessment definitions was partially supported. The well-fitting structural model showed that increased agreement with four different conceptions of assessment

predicted changes in how assessment was defined, as expected, but in a different fashion than was anticipated. Defining assessment as formal test-like practices was predicted by a greater agreement that assessment helped teachers improve teaching and decreased with greater agreement that assessment was personally enjoyable. Whereas, defining assessment as interactive-informal practices was predicted by a perceived increase in classroom benefits and increased ignoring of assessment. The students therefore seemed to associate assessment for learning with definitions of assessment as test-like practices, while conceptions of assessment as helping social environments and being irrelevant led to assessment being defined in terms of interactive informal practices. In other words, we could expect to hear students say something like: "If assessment is controlled by the teacher, I won't necessarily like it, but it will help them teach me better. The interactive assessments are good for class dynamics but I don't need to pay much attention to them". These findings, despite their statistical significance, require confirmation with independent samples and validation through direct examination of students' thinking.

4.1. Interactive-informal assessment practices

While the regression weights to interactive-informal assessment are small, the pattern of weights from the conceptions to both assessment practices combine to form a compelling picture. How these students interpreted interactive-informal assessment practices is potentially a matter of some concern. Assessment reform advocates report increased student achievement through the use of interactive-informal assessment practices. The use of student-centred, alternative assessments (e.g., portfolios, peer-assessment, and self-assessment) and the use of teacher-student interaction (e.g., conferences, checklists, questioning, observation) are claimed to raise student academic performance (Black & Wiliam, 1998; Crooks, 1988). However, this study shows that students did not define assessment with those practices; instead they associated assessment with test-like practices. Furthermore, the students associated informal assessment practices with being affectively beneficial for the class and with being irrelevant. Additionally, Study 2 (Brown & Hirschfeld, in press) clearly associated the same personally enjoyable and irrelevant conceptions of assessment with lowered academic performance. Thus, we have grounds to be concerned that, while teachers are being encouraged to use more informal assessments, students perceive those assessments as being anti-learning.

There are several possible explanations for this disjuncture between practices that teachers are being told are good for academic performance and the conceptions students have of those same practices. It may simply be that New Zealand teachers have not made it clear or persuaded students that participating in these informal and interactive assessment practices helps improve learning. It may be that these practices do not count towards the high-stakes qualifications assessments and are thus disregarded by the students (Bourke, 1996). It may be that the impact of the high-stakes qualifications system that begins in Year 11 is impacting on the assessment practices of Year 9 and 10 students such that the interactive-informal practices are not being implemented (Mizutani, 2006). It may be that the students have potentially legitimate concerns about the reliability and validity of this kind of assessment practice. It may even be that, notwithstanding the rhetoric of assessment reform, it simply isn't happening in New Zealand secondary schools. Alternately, it may be that, for now, the students are right in thinking that teacher-controlled assessment practices lead to the greatest information to the student and the teacher about how to improve

student learning. This last explanation is consistent with Bourke's (2000) reporting that students saw self-assessment as being a matter of relying on confirmation from the teacher and with Peterson and Irving's (in press) study in which peer- and self-assessment were considered irrelevant because only the teacher's comments mattered.

Finally, it may even be that there is a need for meaningful application of both traditional test-like and interactive-informal assessment practices within school classrooms if the highest level of academic performance is to be attained by all. Clearly, further research within this space is required to explore and resolve this tension.

4.2. Implications for classroom practices

New Zealand secondary teachers are faced with a somewhat serious dilemma—only test-like practices were seen by students as improving learning, while interactive informal assessment practices were seen as socially beneficial or to be ignored. A teacher aiming to maximise student academic performance can take advantage of students' commitment to using assessment to improve by implementing practices that maximise information from the teacher to the student about how to improve. That probably means, at least for now, de-emphasising the fun aspects of interactive-informal assessment practices and more clearly linking them to processes that are seen as test-like. If the teacher ensures that the checklists, peer-assessments, self-assessments, and so on, which are associated with increasing student self-regulation, are clearly guided and calibrated against test-based grades, judgements, and feedback, then students may be persuaded that these activities are much more about improving learning than they are about entertaining or occupying class time.

That would mean teachers have to find ways to make such assessment activities actually count towards qualifications and improved student learning. The challenge will be to persuade students that assessment activities other than tests can count, can improve learning, and are reliable. This may also mean persuading teachers of the same requirement. At least at the tertiary level, there are studies that have shown that self- and peer-assessments can be made to count and are reasonably well-correlated with teacher grading (Cho, Schunn, & Wilson, 2006; Falchikov & Boud, 1989; Falchikov & Goldfinch, 2000). Nevertheless, persuading the students of the merits of interactive-informal assessment for learning will probably require a somewhat different approach to that currently being implemented.

4.3. Future research

These structural relations explain a large amount of variance in the teacher-controlled practices and a much smaller amount of interactive-informal practices. Additionally, the practices and beliefs examined in this study are potentially modifiable in classroom contexts, unlike factors such as sex, ethnicity, or socio-economic status and self-regulation studies have shown that modifying student thinking does lead to greater academic performance (Zimmerman, 2001). Greater importance could be given to the results if it were shown that assessment practices combined with conceptions of assessment to predict academic performance; a project we are currently conducting (Brown, Irving, & Peterson, 2007 submitted). Nevertheless, there is much yet to understand. Several of the conceptions of assessment items refer to student learning (e.g., assessment interferes with my learning). Learning is a broad term and it is possible that the students interpreted the

term in different ways when they were answering the questionnaire; a matter we are currently investigating (Peterson, Irving, & Brown, 2007). Further, all our findings are based on data from only two instruments; hence it is possible that factors outside the scope of this study are mediating the associations that we found. Another problem is the choice of level of analysis. We used individual responses as input for our analysis. Additional variance may be explained by taking into account class or school effects in a multi-level modeling approach. Such analysis would also address the issue of whether the results are a reflection of the schools that took part in the study.

This study has shown that the Student Conceptions of Assessment-Version V inventory (SCoA-V) has potential as a research tool; cross-validation evidence with other samples, other instruments, and other data collection procedures is still needed. Further studies are needed to evaluate the inventory's usefulness, as well as applicability in other school systems. For example, mapping of the current model to actual student academic performance is currently being carried out, and other techniques are needed to examine how inventory results relate to how students prepare for and respond to assessment processes. How students practice assessment in their own learning probably mediates how their conceptions of assessment relate to their learning. This study, nevertheless, gives researchers an additional tool in understanding what students think about and how they experience educational assessment.

References

- Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50, 179-211.
- Ajzen, I. (2002). Residual effects of past on later behavior: Habituation and reasoned action perspectives. Personality and Social Psychology Review, 6(2), 107-122.
- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice - a review and recommended 2-step approach. Psychological Bulletin, 103(3), 411-423.
- Arbuckle, J. L. (2005). AMOS (Version 6.0.0). Spring House, PA: Amos Development Corporation.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education, 5(1), 7-74.
- Blaikie, F., Schonau, D., & Steers, J. (2004). Preparing for portfolio assessment in art and design: A study of opinions and experiences of exiting secondary school students in Canada, England, and The Netherlands. The International Journal of Art & Design Education, 23(3), 302-315.
- Bloxham, S., & West, A. (2004). Understanding the rules of the game: Marking peer assessment as a medium for developing students' conception of assessment. Assessment & Evaluation in Higher Education, 39(6), 721-733.
- Bourke, R. (1996, December). But is it right? Students' conceptions of assessment. Paper presented to the Annual Conference of the New Zealand Association for Research in Education, Nelson, New Zealand.
- Bourke, R. (2000). Students' conceptions of learning and self-assessment in context. Unpublished doctoral dissertation, Massey University, Palmerston North, New Zealand.

Brookhart, S. M., & Bronowicz, D. L. (2003). 'I don't like writing. It makes my fingers hurt': Students talk about their classroom assessments. Assessment in Education, 10(2), 221-242.

Brown, G. T. L. (2004). Measuring attitude with positively packed self-report ratings: Comparison of agreement and frequency scales. Psychological Reports, 94, 1015-1024.

Brown, G. T. L. (2006). Secondary school students' conceptions of assessment: A survey of four schools (Conceptions of Assessment and Feedback Project Rep. No. 5). Auckland, NZ: University of Auckland.

Brown, G. T. L., & Hirschfeld, G. H. F. (2005). Secondary school students' conceptions of assessment (Conceptions of Assessment and Feedback Project Rep. No. 4). Auckland: University of Auckland.

Brown, G. T. L., & Hirschfeld, G. H. F. (2007). Students' conceptions of assessment and mathematics: Self-regulation raises achievement. Australian Journal of Educational and Developmental Psychology, 7, 63-74.

Brown, G. T. L., & Hirschfeld, G. H. F. (in press). Students' conceptions of assessment: Links to outcomes. Assessment in Education: Principles, Policy and Practice.

Brown, G. T. L., Irving, S. E., & Peterson, E. R. (2007 submitted). Beliefs that make a difference: Students' conceptions of assessment and academic performance. Paper submitted to the Biannual Conference of the International Test Commission, Liverpool, UK.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling, 9(2), 233-255.

Cho, K., Schunn, C. D., & Wilson, R. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. Journal of Educational Psychology, 98(4), 891-901.

Crooks, T. (1988). The impact of classroom evaluation practices on students. Review of Educational Research, 58(4), 438-481.

Entwistle, N. J., & Peterson, E. R. (2004). Conceptions of learning and knowledge in higher education: Relationships with study behaviour and influences of learning environments. International Journal of Educational Research, 41, 407-428.

Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. Review of Educational Research, 59(4), 395-430.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. Review of Educational Research, 70(3), 287-322.

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. Multivariate Behavioral Research, 42(3), 509-529.

Fodor, J. A. (1998). Concepts: Where Cognitive Science Went Wrong. Oxford: Clarendon Press.

Green, T. F. (1971). The Activities of Teaching. New York: McGraw-Hill.

Gulikers, J. T. M., Kester, L., Kirschner, P. A., & Bastiaens, T. J. (2007). The effect of practical experience on perceptions of assessment authenticity, study approach, and learning outcomes. Learning & Instruction. DOI: 10.1016/j.learninstruc.2007.1002.1012.

Hair, J. F., Jr., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). Multivariate Data Analysis. (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Hirschfeld, G. H. F., & Brown, G. T. L. (under review). Students' conceptions of assessment: Factorial and structural invariance of the SCoA across sex, age and

ethnicity. European Journal of Psychological Assessment,

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6, 1-55.

Klockars, A. J., & Yamagishi, M. (1988). The influence of labels and positions in rating scales. Journal of Educational Measurement, 25(2), 85-96.

Lam, T. C. M., & Klockars, A. J. (1982). Anchor point effects on the equivalence of questionnaire items. Journal of Educational Measurement, 19(4), 317-322.

Laurence, S., & Margolis, E. (1999). Concepts and cognitive science. In E. Margolis & S. Laurence (Eds.), Concepts: Core Readings (pp. 3-81). Cambridge, MA: MIT Press.

Mizutani, S. (2006, July). Understanding washback of NCEA on the teaching and learning of Japanese – Not as straightforward as you think? Paper presented to the New Zealand Association of Language Teachers (NZALT) Conference, Auckland, New Zealand.

Moni, K. B., van Kraayenoord, C. E., & Baker, C. D. (2002). Students' perceptions of literacy assessment. Assessment in Education, 9(3), 319-342.

Olsen, L., & Moore, M. (1984). Voices from the classroom: Students and teachers speak out on the quality of teaching in our schools. Oakland, CA: Students for Quality Teaching Project / Citizens Policy Center.

Osborne, J. W., & Costello, A. B. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical Assessment Research & Evaluation, 10(7), Available online: <http://www.pareonline.net/pdf/v10n17.pdf>.

Pajares, M. F., & Graham, L. (1998). Formalist thinking and language arts instruction: Teachers' and students' beliefs about truth and caring in the teaching conversation. Teaching & Teacher Education, 14(8), 855-870.

Peterson, E. R., Irving, S. E., & Brown, G. T. L. (2007, August). Students' conceptions of learning: Studies of New Zealand secondary students within the Conceptions of Assessment and Feedback Project. Paper presented to the Biannual Conference of the European Association for Research in Learning and Instruction (EARLI), Budapest, Hungary.

Peterson, E. R., & Irving, S. E. (in press). Secondary school students' conceptions of assessment and feedback. Learning and Instruction. DOI:10.1016/j.learninstruc.2007.05.001

Reay, D., & Wiliam, D. (1999). 'I'll be a nothing': Structure, agency and the construction of identity through assessment. British Educational Research Journal, 25(3), 343-354.

Sambell, K., McDowell, L., & Brown, S. (1997). "But is it fair?": An exploratory study of student perceptions of the consequential validity of assessment. Studies in Educational Evaluation, 23(4), 349-371.

Segers, M., & Dochy, F. (2001). New assessment forms in problem-based learning: The value-added of the students' perspective. Studies in Higher Education, 31(5), 327-343.

Steiger, J. H. (2000). Point estimation, hypothesis testing, and interval estimation using the RMSEA: Some comments and a reply to Hayduk and Glaser. Structural Equation Modeling, 7(2), 149-162.

Torrance, H., & Pryor, J. (1998). Investigating formative assessment: Teaching, learning and assessment in the classroom. Buckingham, UK: Open University Press.

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), Handbook of research on mathematics teaching and learning (pp. 127-146). New York: MacMillan.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organizational Research Methods, 3(4), 4-70.

Walpole, M., McDonough, P. M., Bauer, C. J., Gibson, C., Kanyi, K., & Toliver, R. (2005). This test is unfair: Urban African American and Latino high school students' perceptions of standardized college admission tests. Urban Education, 40(3), 321-349.

Weekers, A. M., Veldkamp, B. P. & Brown, G. T. L. (2007 submitted). Analyzing the dimensionality of a students' conceptions of assessment inventory. Paper submitted to the Biannual Conference of the International Test Commission, Liverpool, UK.

Zeidner, M. (1987). Essay versus multiple-choice type classroom exams: The student's perspective. Journal of Educational Research, 80(6), 352-358.

Zeidner, M. (1992). Key facets of classroom grading: A comparison of teacher and student perspectives. Contemporary Educational Psychology, 17, 224-243.

Zimmerman, B. J. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.), Self-regulated learning and academic achievement: Theoretical perspectives (2nd ed., pp. 1-37). Mahwah, NJ: Erlbaum.

Table 1
Second-order, first-order factors and exemplar items of the SCoA-V inventory

| Second-order factors | First-order factors | Items |
|----------------------|---------------------------|---|
| Accountability | School Accountability | Assessment provides information on how well schools are doing Assessment measures the worth or quality of schools *Assessment keeps schools honest and up-to-scratch |
| | Student Accountability | Assessment results show how intelligent I am Assessment is important for my future career or job Assessment is important for my future career or job |
| Affect/Benefit | Class Environment | Assessment encourages my class to work together and help each other Assessment motivates me and my classmates to help each other Our class becomes more supportive when we are assessed |
| | Personal Enjoyment | Assessment is an engaging and enjoyable experience for me I find myself really enjoying learning when I am assessed |
| Improvement | Student Improvement | I make use of the feedback I get to improve my learning I look at what I got wrong or did poorly on to guide what I should learn next I use assessments to take responsibility for my next learning steps |
| | Teacher Improves Learning | Assessment helps teachers track my progress Assessment is a way to determine how much I have learned from teaching My teachers use assessment to help me improve |
| Irrelevance | Bad | Assessment is unfair to students Assessment interferes with my learning Teachers are over-assessing |
| | Ignore | I ignore assessment information I ignore or throw away my assessment results Assessment has little impact on my learning |

* item removed from inventory after analysis.

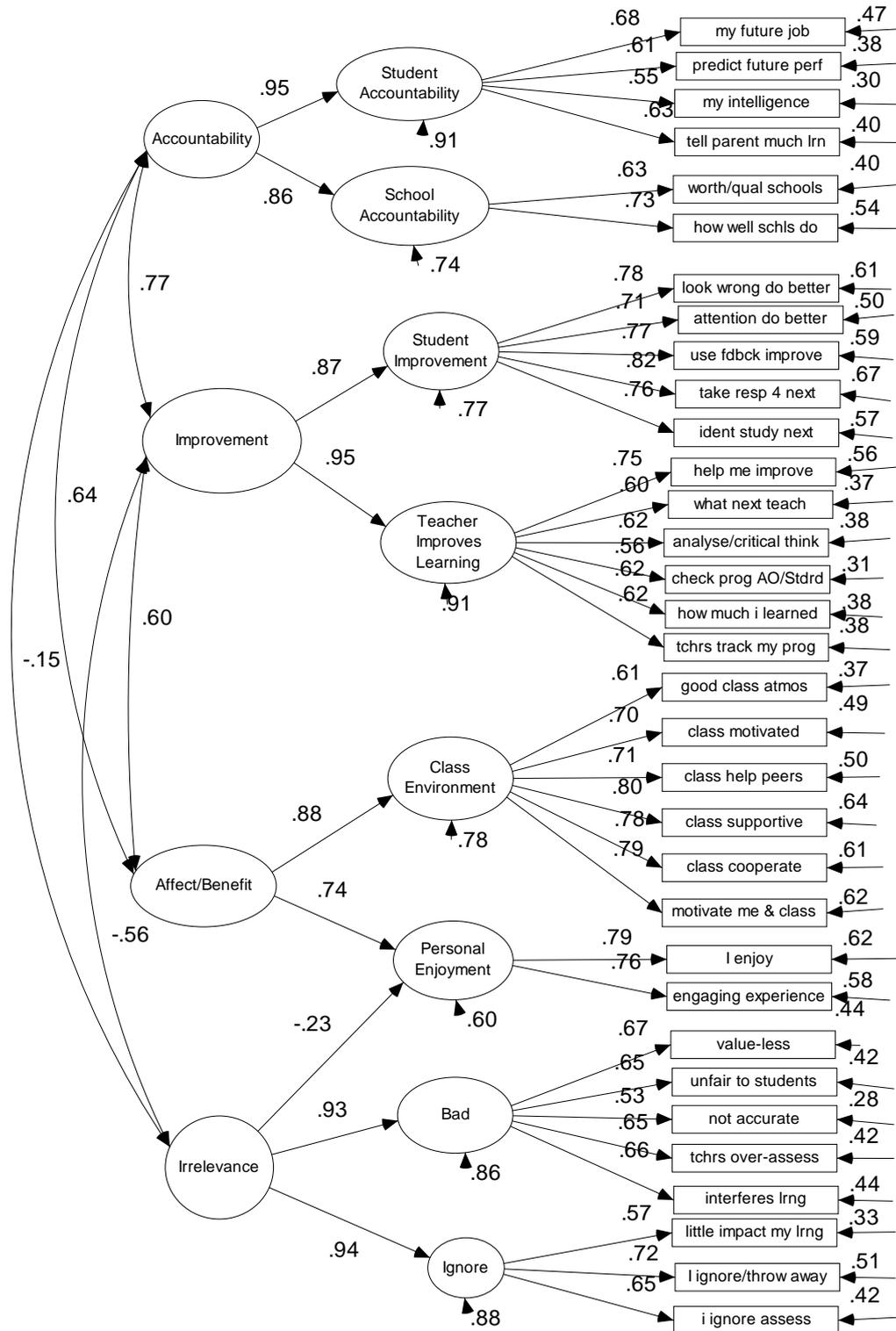


Fig. 1. Students' conceptions of assessment measurement model.

Note. Standardized regression weights shown; SMC shown on residual paths.

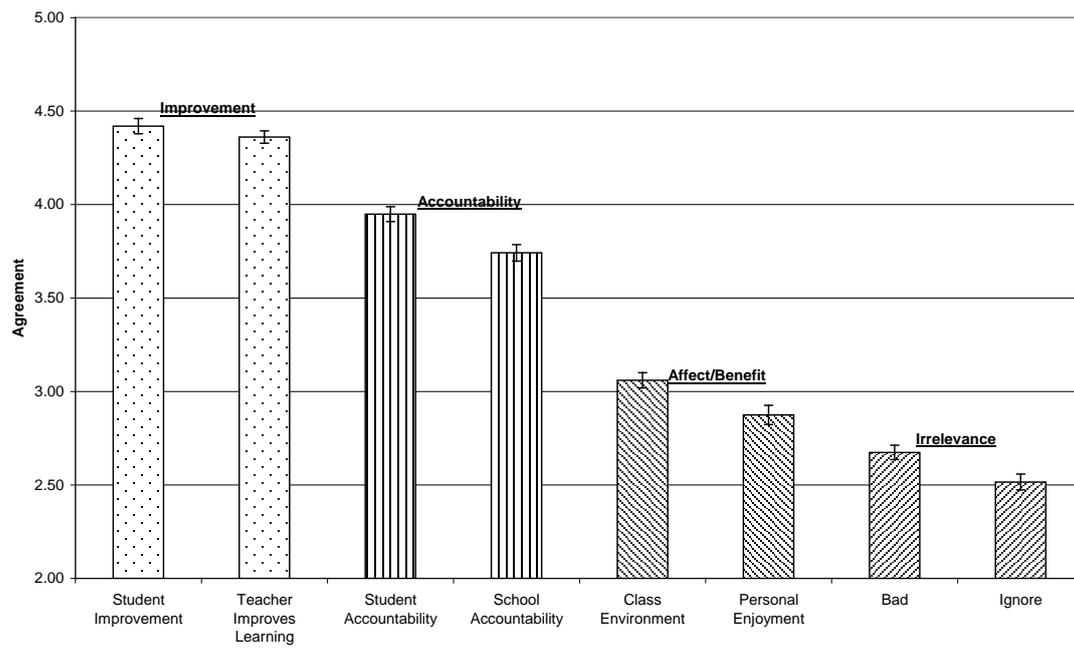


Fig. 2. Strength of agreement per first-order factor.

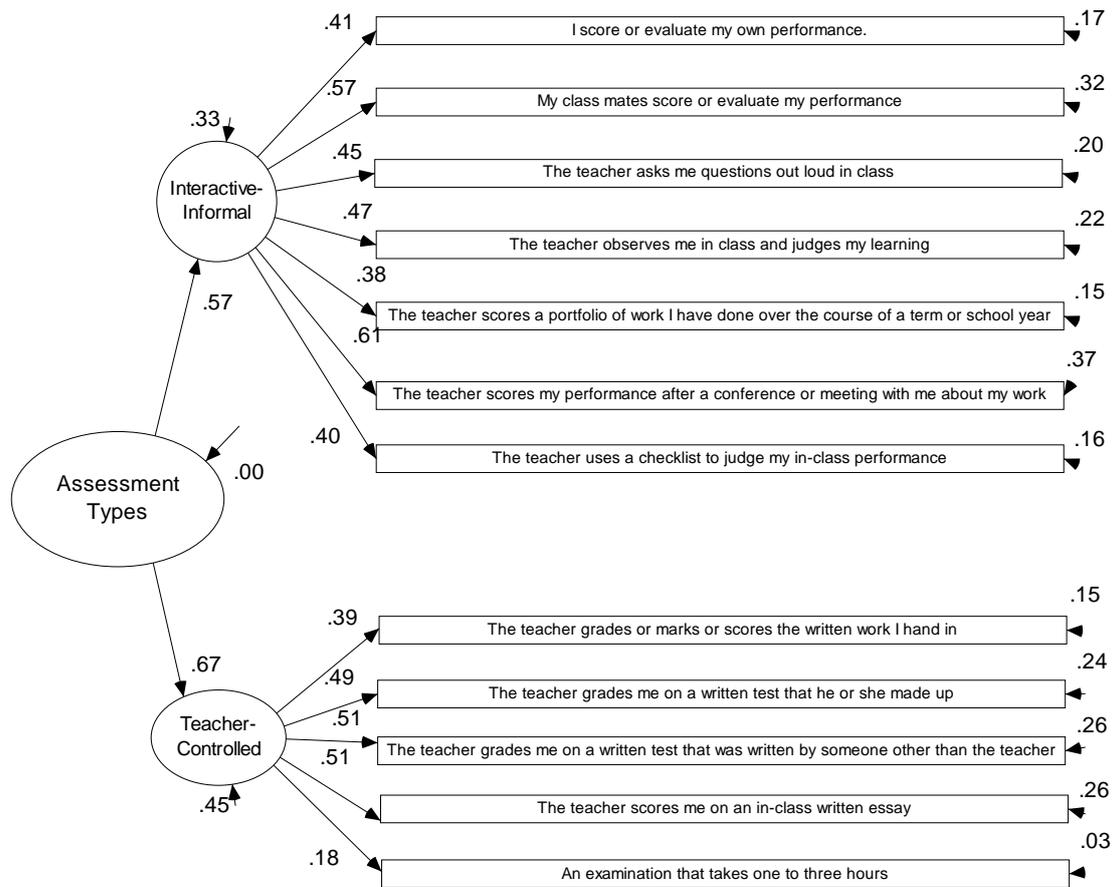


Fig. 3. Assessment definitions measurement model.

Note. Standardized regression weights shown; SMC shown on residual paths.

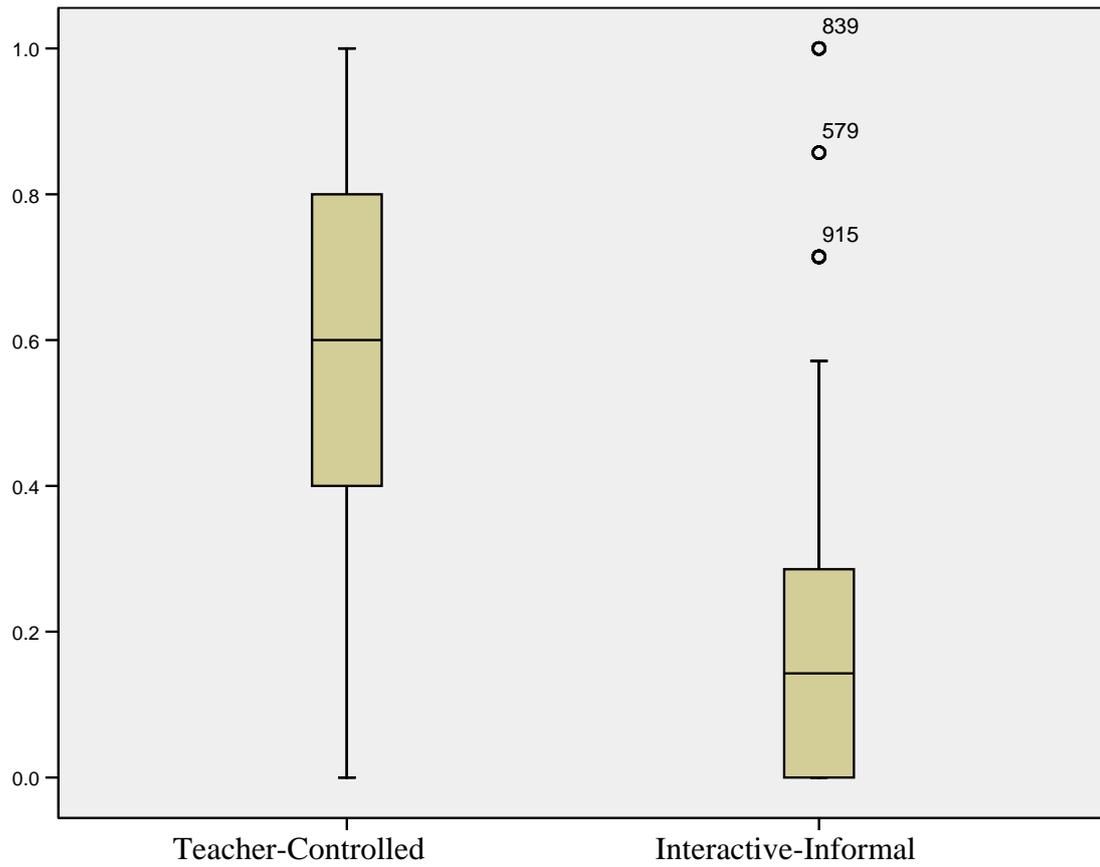


Fig. 4. Score distributions by assessment definition. The vertical axis represents mean number of assessment practices selected by participants for each assessment definition such that 1 = all and 0 = none.

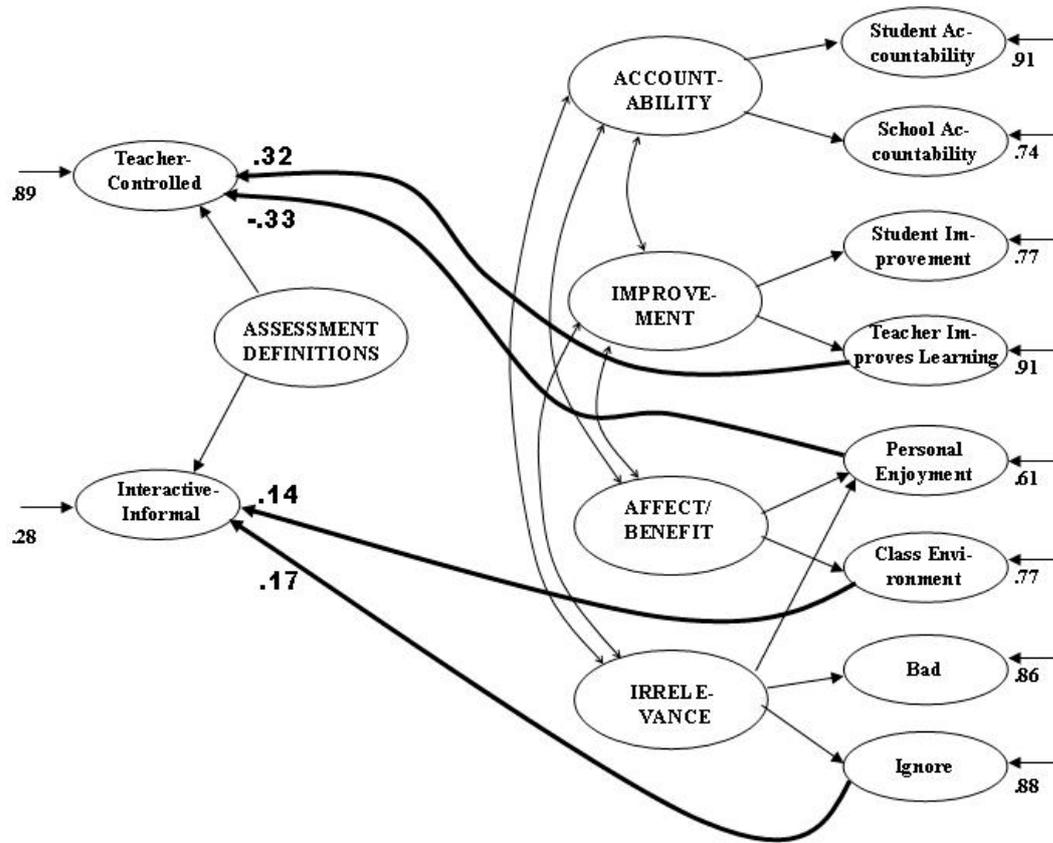


Fig. 5. Students' conceptions of assessment and assessment definitions structural model. Statistically significant structural paths are shown in bold as standardized regression weights. Note. Standardized regression weights shown; SMC shown on residual paths.