Australian Journal of Educational & Developmental Psychology. Vol 7, 2007, pp 63-74

Students' Conceptions of Assessment and Mathematics: Self-Regulation Raises Achievement[1]

Gavin T. L. Brown[2]
University of Auckland,
&
Gerrit H. F. Hirschfeld
University of Muenster

ABSTRACT

A study of New Zealand secondary school students using four self-report inventories of conceptions of assessment found four robust independent measurement models. Four structural models mapped the conceptions of assessment to mathematics achievement taking into account student ethnicity and student sex. The conceptions that assessment makes students accountable and was beneficial for students loaded positively on achievement, while the conceptions that assessment is fun and assessment is ignored had negative loadings on achievement. These findings are consistent with self-regulation theory and formative assessment, suggesting that students who use assessment to take responsibility for their learning by using assessment formatively will attain increased mathematics outcomes.

Keywords: Assessment, secondary school students, conceptions, mathematics

INTRODUCTION

A conception is a mental construct or representation of reality (Kelly, 1991), communicated in language or metaphors (Fodor, 1998; Lakoff & Johnson, 2003), containing beliefs, meanings, preferences, and attitudes (Thompson, 1992) and which explains complex and difficult categories of experience (White, 1994) such as assessment. People's purposes towards phenomena are expressed in their conceptions of the phenomena (Fodor, 1998); for example, the concept that assessment ought to be 'formative' rather than 'summative' generally implies assessment should be used to improve teaching and learning and not give students final grades or scores. The nature of these mental representations is contested (Laurence & Margolis, 1999), however, it would appear that our conceptions are 'in-pieces' (diSessa, 1988) or 'informationally atomistic' (Fodor, 1998) or in clusters (Green, 1971).

[1] **Acknowledgement:** This research was conducted in conjunction with the Assessment Tools for Teaching and Learning (asTTle) Project (www.asttle.org.nz) at The University of Auckland, with ethical approval from the UoA Human Participants Ethics Committee Reference 2000/296. The cooperation of the schools, teachers, and their students participating in the project is acknowledged. The work of the asTTle data team, led by Andrea MacKay, in preparing the data is especially appreciated.

[2] **Contact**
Dr. Gavin T L Brown,
School of Teaching, Learning and Development,
Faculty of Education, University of Auckland,
Private Bag 92019, Auckland, New Zealand.
Email: gt.brown@auckland.ac.nz

What this means is that people may simultaneously hold multiple, and possibly even contradictory, conceptions of a phenomenon without being disturbed by such contradiction—a result reported by Brown's (2004a) investigation of New Zealand teachers' agreed with both assessment for improvement and assessment for school accountability. Further, it implies that a single conception may actually serve multiple purposes without threatening the integrity of the conception itself.

Students' thinking about educational processes is important because there is evidence that how they understand those processes impacts on their educational experiences. Students who take responsibility for their learning generally achieve more (Reeve, Bolt, & Cai, 1999; Ryan & Grolnick, 1986); whereas, those who locate control or apportion responsibility elsewhere (Rotter, 1982) or who lack confidence to achieve (Bandura, 1989; Pajares, 1996) tend to achieve less. In higher education, students' learning is more influenced by their perceptions of the educational environment then by the actual educational practises (Entwistle, 1991). Furthermore, students' conceptions of assessment are of particular importance because assessment has a significant impact on the quality of learning (Entwistle & Entwistle, 1991; Marton and Säljö, 1997; Ramsden, 1997). Pajares (1992) has argued that teachers' conceptions of educational processes are a product of their educational experiences as students, suggesting strongly that similar conceptions might be found in both teachers and students. However, how secondary school students conceive of assessment and how those conceptions relate to their academic performances is much less understood.

Secondary School Students' Conceptions of Assessment.

At secondary school multiple and conflicting conceptions of assessment are evident. Zeidner (1992) reported that Israeli junior and senior high school students, when offered the choice of four different purposes of assessment (i.e., summarising student achievement, arousing student interest and motivation, evaluating quality of teaching, or administrative) had a stronger perception of assessment as evaluating student achievement than for improving motivation or behaviour. 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".

Within the context of student accountability (i.e., contributing to course grades), it is not surprising to find evidence that students preferred assessment methods that that they perceived as maximising their grades or, more charitably, improving their learning. For example, lower performing and more test-anxious students preferred multiple-choice formats (Zeidner, 1987). However, there is more evidence that students became increasingly negative towards assessment as the consequences for their lives increased. For example, Australian students in their first year of high school became increasingly negative to assessment because of the increased volume of it compared to primary school and because of the perceived subjectivity of teacher assessment decisions (Moni, van Kraayenoord, & Baker, 2002). Urban African American and Latino high school seniors also perceived the high-stakes university entrance tests as an unfair obstacle (partly because of its one-shot nature) upon their life chances (Walpole, McDonough, Bauer, Gibson, Kanyi, & Toliver, 2005), though in contrast to the English children (Reay & Wiliam, 1999), the tests rather than themselves were blamed for poor results. In a low-stakes context, New Zealand high school students generally enjoyed doing standardised tests that had a mixture of multiple-choice and open-ended item formats (Hattie, Brown, Ward, Irving, & Keegan, 2006). However, no meaningful correlations were found between enjoyment of the assessments with achievement in reading, mathematics, pānui, and tuhituhi (mean correlation over ten evaluation factors and four subjects was $r = .013$; $SD = .11$; all statistically significant).

There is evidence that students simply prefer whatever system of assessment that they experience, regardless of the merits or deficiencies of that system (Blaikies, Schönau, & Steers, 2004); thus, students may not really be in a position to evaluate assessment methods that are not in their experience. Perhaps, students simply normalise and adopt the values,

beliefs, and preferences of their teachers (Zeidner, 1992). Given this habit and the doubt as to the relationship of student preference for format to outcomes, it may be that the format of assessment itself is irrelevant to student achievement and that we should be examining students' perceptions of assessment around the purposes of assessment.

There is evidence from Brown and Hirschfeld (accepted) that students' conceptions of the purpose of assessment are related to differing degrees of achievement. Results, from about 3500 students, showed four conceptions of assessment (assessment makes students accountable, assessment makes schools accountable, assessment is fun, and assessment is ignored) and these conceptions were meaningfully related to student achievement in reading (χ^2 = 803.521; *df* = 81; RMSEA = .051; TLI = .91; CFI = .93). It was found that the conception that assessment was about making students accountable loaded positively on reading achievement scores (λ = .42), whereas, the conceptions that assessment was fun (λ = -.24), assessment was ignored (λ = -.14), and assessment makes schools accountable (λ = -.27) all had negative loadings on reading achievement scores. These results were interpreted in the light of self-regulation theory (Zimmerman, 2001) which states that students who see assessment as a constructive force for personal responsibility gain higher grades, while those who seek to 'blame' schools or teacher for assessment results, those who do not take assessment seriously, or who ignore assessment receive lower grades.

To summarise, it would appear that at least three major conceptions of assessment among secondary students can be inferred:

- assessment is a negative thing because it is unfair, bad, or interfering to the students' learning,
- assessment, including classroom assessment, acts to make students themselves as well as their teachers and schools accountable, and
- assessment, or at least some formats or procedures, may be beneficial, even enjoyable, in improving the quality of student learning.

METHOD

This paper reports the results of a pilot study into students' conceptions of assessment as they relate to mathematics achievement. It examines the strength of students' agreement with different purposes of assessment and links their conceptions of assessment to achievement outcomes on standardised national assessments of mathematics. In this way we come closer to understanding whether certain conceptions towards assessment have any relationship to increased learning outcomes. The research design used self-report survey questionnaire responses and exploratory and confirmatory factor analyses identify the conceptions students have, how those conceptions related to each other, and how they related to academic outcomes. The next section describes the instrumentation for measuring students' conceptions of assessment and academic achievement; followed by a description of the analytic procedures used to evaluate student self-reports. Subsequently, the results are provided and a conclusion provides a summary of main points with implications for practice and further research.

Instruments

A self-report questionnaire survey was used to elicit students' conceptions of assessment and student responses to the Assessment Tools for Teaching and Learning curriculum-based knowledge and skill in mathematics tests were used to determine learning outcomes.

Students' Conceptions of Assessment

Previous research into teachers' conceptions of assessment has shown the importance of the purpose of assessment as a way of understanding what assessment means (Brown, 2002; 2004a; 2006). Thus, in addition to the literature reviewed, items that had contributed to identifying teachers' conceptions of assessment were also used in developing a conceptions of assessment inventory for students. This is probably a legitimate expectation since there is a

strong likelihood that teachers learned their conceptions of assessment partly from their experiences of being school students (Pajares, 1992).

In this study, 49 items for the student questionnaire (Students' Conceptions of Assessment V1—SCoA-I) were designed to map to four main concepts: assessment makes schools or students accountable (13 items), assessment improves teaching and learning (13 items), assessment is negative or bad (13 items), and assessment provides a useful description of performance (11 items). Wording changes were made for items adopted from the Teachers' Conceptions of Assessment inventory to make them read from the student perspective; for example, the word "teacher" was replaced by "students". The items were presented in four forms (Maths A to Maths D) at the end of a 40 minute mathematics test. Because of very limited time available and to increase student completion of the item sets, each form had only 11 to 13 items relating to one conception only. Thus, analysis of each conception was done independent of the other factors, but all results were still structurally linked to mathematics achievement.

Adjectives in balanced response formats (e.g., Likert scales) are simple mirrors of each other; that is, strongly disagree and disagree are balanced with agree and strongly agree. One of the weaknesses of this format, not unexpectedly, is that if most or all participants are inclined to be positive about the object being evaluated, having only two shades of positive orientation results in lowered variance in the responses. Variance is a prerequisite of accurate measurement. It is likely that balanced response anchors will not provide variance when participants are inclined to respond positively to all items because they are deemed equally true or valuable. Thus, skewed response formats have been tried and found effective in increasing variance when respondents have opinions that are skewed towards one end or the other of the response scale (Brown, 2004b; Klockars & Yamagishi, 1988; Lam & Klockars, 1982). Scales that have more positive response options are positively-packed, while those with more negative options are negatively-packed. When using positively or negatively packed response formats some care is needed in selecting the intermediate adjectives. Hattie (personal communication, February, 1999) reported unpublished research (similar in method to that of Lam & Klockars, 1982) which indicated that the following adjectives would provide nearly-equal intervals on an underlying scale of agreement (i.e., strongly agree, mostly agree, moderately agree, slightly agree, mostly disagree, and strongly disagree). This rating scale has been found to generate well-fitting data and good variance in a students' conceptions of learning questionnaire (Brown, 2004b) and in a teachers' conceptions of assessment questionnaire (Brown, 2004a).

Learning Outcomes

The outcome measures were secondary school students' performance on the Assessment Tools for Teaching and Learning (asTTle) mathematics tests. The asTTle Project developed, under government contract, a bank of standardised assessment items for reading, mathematics, and writing calibrated against New Zealand curriculum levels and norms (Hattie, Brown, & Keegan, 2003). The asTTle software reports student achievement in each subject using an item response theory (IRT) calibrated scale score (Hattie, Brown, Keegan, MacKay, Irving, Cutforth, et al., 2004). These assessments contained both multiple-choice and open-ended (though brief) response format items and participation was voluntary. The only use made of the assessment results was to calibrate the item psychometric characteristics and establish national norms for performance at each year level. Each asTTle mathematics test generates scaled scores for total score, curriculum content scores, and cognitive processing scores. The four different SCoA forms were attached to the end of four different asTTle mathematics tests administered in 2003 to secondary students.

The asTTle mathematics tests assess eight domains of mathematics knowledge (i.e., number knowledge, number operations, algebra, geometric knowledge, geometric operations, measurement, probability, and statistics). Each test had items covering a random selection of those content areas targeted to the expected ability range for the year group being tested. This meant that the tests covered different mixtures of specific content, but their content was unidimensional at the level of being mathematics. Since the SCoA forms were assigned to

different mathematics tests, the only comparable outcome measures was the total score, which was obtained through one parameter logistic (1PL) IRT[3] calibration of all items and all respondents (Hattie, Brown, Keegan, et al., 2004). Although IRT modelling can include item difficulty, discrimination, and pseudo-chance parameters (Embretson & Reise, 2000), the student's ability in mathematics was determined by a 1PL formula that took into account the difficulty of each item answered correctly by the student regardless of the mixture of items faced by each participant. Thus, this study was able to compare students' conceptions of assessment to their mathematics ability, regardless of the different items presented to each group of students.

Analyses

All participants who had answered less than 90% of presented conceptions items were dropped from the analysis. With the balance, data missing at random was imputed using the SPSS expectation maximisation (EM) missing values procedure (Hair, Anderson, Tatham, & Black, 1998). Inspection of means and standard deviations indicated that the EM procedure caused only minimal difference to the data (i.e., differences noticeable only at the .01 level).

The conceptions of assessment response categories were scored 1 for the most negative and 6 for the most positive. Note that the asTTle mathematics scores used in the analysis were the IRT logit scores ($M = 4.29$, $SD = 1.61$), rather than the standardised linear transformation aMs scores (sample $M = 700$, $SD = 100$) since there were problems with the estimated path coefficients and covariances attributable to the vastly different scales (Kim & Mueller, 1978).

Each analysis was carried out in three steps. First, the internal structure of the item-sets was explored using maximum likelihood factor analysis with oblique rotation (Osborne & Costello, 2005). Items that had poor fit characteristics were identified and dropped from subsequent analyses; this poor fit included items with loadings below .30, those with cross-loadings greater than .30 on another factor, and those that had poor theoretical fit with the other items in the factor. The second and third steps were similar to Anderson's and Gerbing's (1988) two-step analysis in that the measurement models were tested first before analysing the structural relations to achievement. Accordingly, the second step used maximum likelihood confirmatory factor analysis to validate the factor structure of the measurement model. Solutions that had reasonable fit characteristics (e.g., CFI or TLI > .90, RMSEA <.08) were utilised in subsequent analyses. It is noted that such analyses are most robust when sample size is greater than 500 (Chou & Bentler, 1995), which applies to only one of the four conceptions reported here. CFA was conducted with AMOS (Arbuckle, 2003).

Third, structural models were constructed by including outcome measures (i.e., asTTle mathematics achievement) and student demographic variables (i.e., student sex and student ethnicity) into the model. Student sex was used in the structural model by treating Female as 0 and Male as 1. Students reported their ethnicity using four major categories—New Zealand European or Pakeha was treated as 1, Maori was 2, Pasifika was 3, and Asian was 4. Thus, negative structural paths from these variables to mathematics ability mean that as sex became male and as ethnicity became non-minority achievement would go down; positive structural paths mean that females and majority ethnicity students would have lower achievement. Note that structural model weights are interpreted as standardised partial regression weights such that the value 1.0 indicates an increase of one standard deviation in the independent variable would cause a one standard deviation increase in the dependent variable. Remember that four structural models were tested because different students responded to different sets of conceptions of assessment items.

[3] Although two-parameter modeling was used to select items for inclusion in the asTTle bank, one-parameter modeling is used to calculate student location scores. Bookmark standard setting procedures were used to map item locations and student ability scores to curriculum levels.

RESULTS

Participants were dropped for having more than 10% missing responses. Thus, the four measurement models made use of valid responses from 1234 secondary school students in Years 9 to 12. Because some three percent did not specify their ethnicity, the structural model analyses had a total of 1191 participants. Thus, the four structural models were based on 162 responses for the accountability conception (Form A), 219 responses for the improvement conception (Form B), 502 for the negative conception (Form C), and 308 for the useful conception (Form D). The recommended sample size for this type of analysis is 500 (Chou & Bentler, 1995), and so except for Form C, all results will be subject to chance artefacts due to sample size.

Given the pilot nature of this study, representative sampling was not considered necessary, nor achieved. This sample was 61% female compared to the asTTle norming population of 49% female, 68% were of New Zealand European ethnicity compared to the asTTle population who were 43.3% New Zealand European, only 14% were Maori compared to 29% in the asTTle population. Thus, the sample was skewed by having too many females and New Zealand/European students, and insufficient Maori students.

Since generalisation to specific sub-populations was not intended, these samples were sufficiently large to give an initial indication of the types of conceptions held by students and how those conceptions might relate to achievement in mathematics. The samples were also sufficiently large to indicate items most likely to be deficient.

Form A: Assessment Makes Students and Schools Accountable

After deleting seven items for poor fit characteristics, six items captured two inter-correlated conceptions of accountability in an acceptable measurement model ($\chi^2 = 15.568$; $df = 8$; RMSEA = .075; TLI = .97). The conception that assessment makes students accountable was based on three items and, likewise, the conception that assessment makes schools accountable.

The structural model consisted of the measurement model plus the asTTle achievement score and the demographic variables of sex and ethnicity, in which the regression weights were freely estimated. The structural model had acceptable fit characteristics ($\chi^2 = 46.561$; $df = 25$; RMSEA = .073; TLI = .94) (Figure 1). On average, the students slightly agreed that assessment made schools accountable ($M = 3.14$; $SD = 1.25$) and agreed moderately that assessment made students accountable ($M = 3.86$; $SD = 1.20$), with the latter conception having small partial regression weight on achievement ($\lambda = .14$). Sex and ethnicity had much stronger predictive relationships to achievement, with increasing performance associated with female sex and New Zealand European ethnicity.

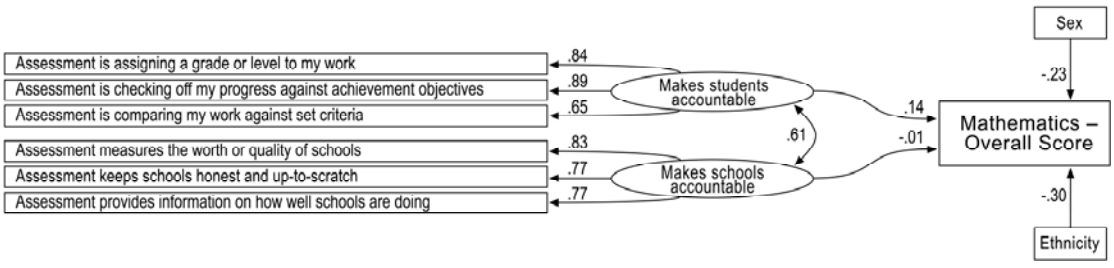

Figure 1. Structural model of students' accountability conceptions of assessment, demographic variables, and mathematics achievement in Form A.

Together, the two variables accounted for 3.9 % of variance in the mathematics score not accounted for by sex or ethnicity; this is an effect size of f^2=.04, a trivial effect (Cohen, 1992). There existed, with this sample of 162 students, a small relationship between students using assessment to make themselves accountable for learning and their mathematics scores increasing.

Form B: Assessment Improves Learning

Of the original 13 improvement items, three factors using nine items were found and combined into an acceptable measurement model (χ^2 = 56.099; *df* = 24; RMSEA = .078; TLI = .94). The factors were assessment improves teaching (two items), assessment is good for me (five items), and assessment is fun (two items). The assessment is good for me has strong elements of self-regulation and self-responsibility embedded in it. The correlations between the assessment improves teaching and assessment is good for me and assessment is fun factors were moderate, while the correlation between the assessment is fun and assessment is good for me factors was strong. The students slightly disagreed with the idea that assessment is fun (*M* = 2.61; *SD* =1.32), slightly agreed that it improves teaching (*M* = 3.09; *SD* =1.30), and moderately agreed that it is good for them (*M* = 3.75; *SD* =1.23).

The indices for the structural model were also acceptable (χ^2 = 88.110; *df* = 49; RMSEA = .061; TLI = .94) (Figure 2). The conception that assessment is fun was statistically significant and negatively related to achievement (λ = -.40) while the conception assessment is good for me was statistically significant and positively related to achievement (λ = .55). In contrast, the conception assessment improves teaching was weakly, though not statistically significant, related to achievement. The significance of sex and ethnicity on achievement was much lower than two of these conceptions; the negative paths showing again that increasing scores were associated with female sex and New Zealand European ethnicity.

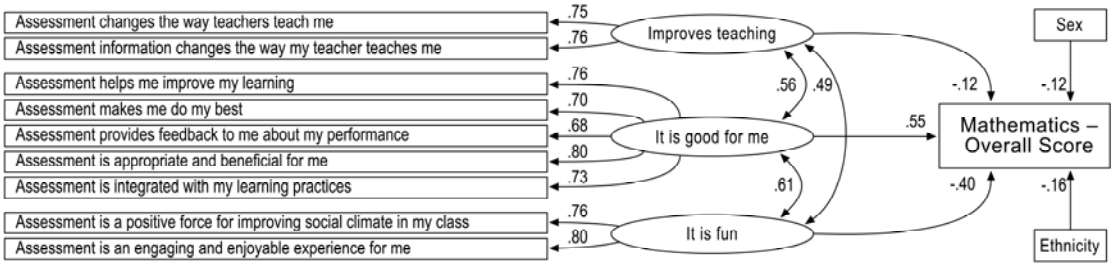

Figure 2. Structural model of students' improvement conceptions of assessment, demographic variables, and mathematics achievement in Form B.

Together, these conceptions explained 6.6% of the variance on top of the demographic variables; this is an effect size of f^2=.07, an effect half-way between trivial and moderate (Cohen, 1992). In other words, there existed, with a sample of 219 students, a robust, albeit small, relationship between students believing assessment was good for them and their mathematics scores increasing. In contrast, the conception that assessment is fun was inversely related to mathematics achievement.

Form C: Assessment is Negative

Eleven of the 13 negative items generated a three factor solution with acceptable fit statistics (χ^2 = 138.612; *df* = 41; RMSEA = .069; TLI = .94): assessment interferes with learning (six items), I ignore assessment (three items), and assessment has error (two items). It is expected that ignoring assessment and believing that it interferes are the obverse side of the self-responsibility coin—students who self-regulate do not ignore assessment, nor do they believe it interferes with learning. The correlations between these three factors ranged from weak to moderate. Students tended to reject the conceptions that assessment interferes with their learning (*M* = 2.78; *SD* =1.01) and that they ignore assessment results (*M* = 2.64; *SD* =1.28), but they agreed slightly that assessment has error (*M* = 3.38; *SD* = 1.15).

The structural model had acceptable fit to the data (χ^2 = 210.441; *df* = 72; RMSEA = .062; TLI = .92) (Figure 3). The assessment interferes with learning conceptions had a statistically significant, negative loading on overall mathematics score (λ = -.21). The other two conceptions had non-statistically significant and weak loadings on the mathematics score. As earlier, male sex and non-New Zealand European ethnicity had weak but negative path weights on achievement.

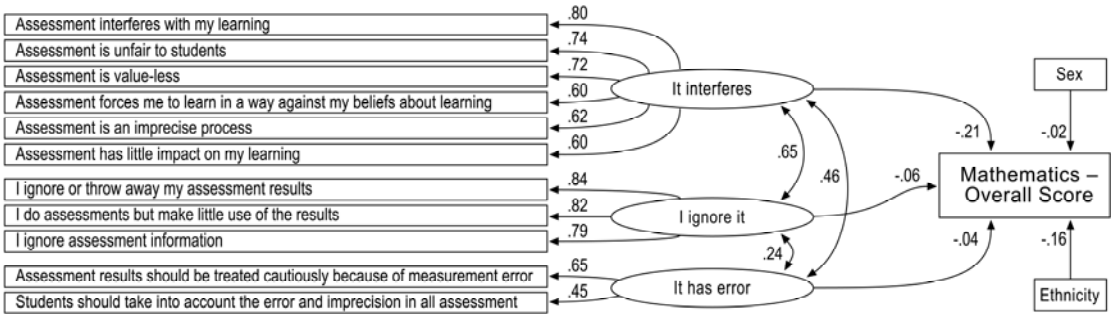

Figure 3. Structural model of students' negative conceptions of assessment, demographic variables, and mathematics achievement in Form C.

Together these three conceptions accounted for 5.5% of the variance of the achievement score; this is an effect size of f^2=.06, a very small effect (Cohen, 1992).. In other words, there existed with a sample of 502 students, a small but statistically significant inverse relationship between conceptions and mathematics performance; mathematics scores decreased as students believed assessment interfered with learning.

Form D: Assessment is Useful

Nine of the 11 useful items were kept in a well fitting measurement model (χ^2 = 50.549; *df* = 24; RMSEA = .060; TLI = .96) with three factors: assessment is valid (four items), assessment captures my thinking (two items), and assessment is reliable (three items). The factors were strongly inter-correlated. Taken together the students moderately agreed that assessment is valid (*M* = 3.70; *SD* =1.06) and slightly agreed that it both captures their thinking (*M* = 3.24; *SD* =1.21) and is objective (*M* = 3.14; *SD* =1.07).

The fit indices for the structural model were good (χ^2 = 96.563; *df* = 49; RMSEA = .056; TLI = .94) and all structural paths were statistically significant at alpha .01 (Figure 4). The conception of validity was positively related to achievement while conceptions of reliability and assessment captures thinking were negatively related to achievement. As before, the path weights from sex and ethnicity were negative (i.e., females and New Zealand Europeans do better) but both were very weak.

Together these conceptions explained 5.3% of the variance in the mathematics score; this is an effect size of f^2=.06, a very small effect (Cohen, 1992). In other words, there existed, with a sample of 308 students, a small, statistically significant, relationship between their conceptions around the validity of assessment and their mathematics achievement.

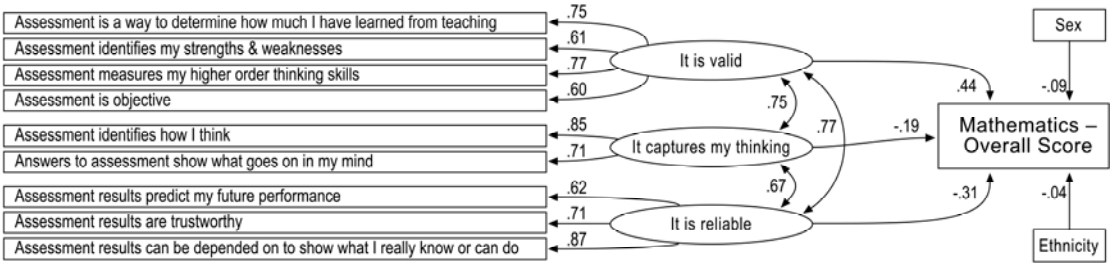

Figure 4. Structural model of students' validity conceptions of assessment, demographic variables, and mathematics achievement in Form D.

DISCUSSION

Four independent measurement models, each made up of two or three inter-correlated first-order factors, concerning students' conceptions of assessment were identified and acceptable fit statistics to the data were found. Four structural models showed that there were small, but non-chance, regression weights between the mathematics achievement scores and four of the conception of assessment. Higher overall mathematics achievement was found among students who conceived that assessment makes students themselves accountable for learning and who conceived of assessment as good for them. In contrast, the more students agreed that assessment interferes with learning or assessment is ignored the lower their mathematics achievement. Thus, it seemed in mathematics that students who conceived of assessment as a means of taking responsibility for learning, who did not ignore assessment, and who used it to improve their learning, tended to get higher achievement scores.

These results are consistent with self-regulation and formative assessment theories (Black & Wiliam, 1998; Crooks, 1988; Zimmerman, 2001). Self-regulation theory states that students who control (e.g., take responsibility, do not blame others, make pro-active use of feedback) their own learning achieve more—and the evidence here is that mathematics scores increase if students agree that assessment itself makes students accountable for learning and if assessment is seen as a beneficial process. Formative assessment focuses on students' active engagement in self-assessment which leads to greater learning outcomes—the students who claim not to ignore or treat it as interfering with their learning achieve more than those who do not use assessment to improve their learning.

However, since these structural models were derived in isolation from each other, this interpretation is somewhat tentative, subject to confirmation from a data set in which it is possible to see how student responses to all conceptions simultaneously relate to their achievement. Additionally, since the effects are quite small, further research with larger samples of students and larger sets of items is needed to ascertain whether students' conceptions of assessment have a meaningful relationship to their academic performance. Another dimension worth exploring is whether these conceptions have meaningful relationships with the various major content areas of mathematics rather than just with mathematics as a whole. To test this possibility students would have to be given tests with parallel content, however, at this time we can see meaningful relationships with mathematics as a whole. This research has also shown that, as might be expected (Satherley, 2006), female sex and New Zealand European ethnicity were most associated with higher mathematics scores, though, the unique contribution of these on academic performance was not as large as might be anticipated. Other demographic or background variables in the students' lives do play a significant role in academic performance, especially socio-economic resources, and future research should seek to examine such effects along with conceptions of assessment.

Nevertheless, these results confirm that students have multiple conceptions of assessment which appeared to be internally consistent rather than contradictory. Further, the results suggest that some conceptions are actually more productive and effective than others in terms of measurable learning outcomes. This is a step towards helping both teachers and students understand that learning outcomes can be enhanced if students treat assessment in a self-regulating and formative manner.

REFERENCES

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. (2003). AMOS (Version 5) [Computer software]. Chicago, IL: SmallWaters Corp.

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist, 44*(9), 1175-1184.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74.

Blaikie, F., Schönau, 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.

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. (2002). *Teachers' Conceptions of Assessment.* Unpublished Doctoral Dissertation, University of Auckland, Auckland, NZ.

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

Brown, G. T. L. (2006). Teachers' conceptions of assessment: Validation of an abridged instrument. *Psychological Reports, 99*, 166-170.

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

Chou, C.-P., & Bentler, P. M. (1995). Estimates and tests in structural equation modeling. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 37-55). Thousand Oaks, CA: Sage.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

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

diSessa, A. A. (1988). Knowledge in pieces. In G. Forman & P. B. Pufall (Eds.), *Constructivism in the Computer Age* (pp. 49–101). Hillsdale, NJ: LEA.

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: LEA.

Entwistle, N. J. (1991). Approaches to learning and perceptions of the learning environment: Introduction to the special issue. *Higher Education, 22*, 201-204.

Entwistle, N. J., & Entwistle, A. (1991). Contrasting forms of understanding for degree examinations: The student experience and its implications. *Higher Education, 22*, 205-227.

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.

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.

Hattie, J. A. C., Brown, G. T. L., & Keegan, P. J. (2003). A national teacher-managed, curriculum-based assessment system: Assessment Tools for Teaching & Learning (asTTle). *International Journal of Learning, 10,* 771-778.

Hattie, J. A. C., Brown, G. T. L., Keegan, P. J., MacKay, A. J., Irving, S. E., Cutforth, S., et al. (2004). *Assessment Tools for Teaching and Learning (asTTle) Version 4, 2005:*

Manual. Wellington, NZ: University of Auckland/ Ministry of Education/ Learning Media.

Hattie, J. A., Brown, G. T. L., Ward, L., Irving, S. E., & Keegan, P. J. (2006). Formative evaluation of an educational assessment technology innovation: Developers' insights into Assessment Tools for Teaching and Learning (asTTle). *Journal of Multi-Disciplinary Evaluation, 5*. Available: http://evaluation.wmich.edu/JMDE/content/JMDE005content/PDFs_JMDE_005/Formative_Evaluation_of_an_Educational_Assessment_Technology_Innovation_Developers_Insights_into_Assessment_Tools_for_Teaching_and_Learning_asTTle.pdf

Kelly, G. A. (1991). *The psychology of personal constructs: A theory of personality* (Vol. 1). London: Routledge.

Kim, J.-O., & Mueller, C. W. (1978). *Factor Analysis: Statistical methods and practical issues* (Vol. 14). Thousand Oaks, CA: Sage

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

Lakoff, G., & Johnson, M. (2003). *Metaphors We Live By*. Chicago, IL: Chicago University Press.

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.

Marton, F., & Säljö, R. (1997). Approaches to learning. In F. Marton, D. Hounsell & N. Entwistle (Eds.), *The Experience of Learning: Implications for Teaching and Studying in Higher Education* (2nd ed., pp. 39-59). Edinburgh: Scottish Academic Press.

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

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). Retrieved July 15, 2005, from http://pareonline.net/getvn.asp?v=1-&n=7

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*, 307-332.

Pajares, M. F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*(4), 543-578.

Ramsden, P. (1997). The context of learning in academic departments. In F. Marton, D. Hounsell & N. Entwistle (Eds.), *The Experience of Learning: Implications for Teaching and Studying in Higher Education* (2nd ed., pp. 198-217). Edinburgh: Scottish Academic Press.

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.

Reeve, J., Bolt, E., & Cai, Y. (1999). Autonomy supportive teachers: How they teach and motivate students. *Journal of Educational Psychology, 91,* 537–548.

Rotter, J. B. (1982). Social learning theory. In N. T. Feather (Ed.). *Expectations and actions: Expectancy-value models in psychology* (pp. 241–260). Hillsdale, NJ: Erlbaum.

Ryan, R. M., & Grolnick, W. S. (1986). Origins and pawns in the classroom. Self-report and projective assessments of individual differences in children's perceptions. *Journal of Personality and Social Psychology, 50,* 550–558.

Satherley, P. (2006). *Student Outcome Overview 2001-2005: Research Findings on Student Achievement in Reading, Writing and Mathematics in New Zealand Schools.* Wellington, NZ: Ministry of Education, Research Division.

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.

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.

White, R. T. (1994). Commentary: Conceptual and conceptional change. *Learning and Instruction, 4*, 117-121.

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. 1–36). Mahwah, NJ: LEA.

Biographical Notes.

Dr Brown's research interests relate to teachers' and students' conceptions, practices, and results around educational assessment. Previously, he was senior project manager of the asTTle team and an assessment researcher at NZCER after being a secondary school teacher for 10 years.

Gerrit Hirschfeld completed a BA Honours thesis in psychology at the University of Muenster and is currently enrolled as a PhD candidate there studying language-production- with EEG. He conducted the data analysis for this paper as part of his BA internship at The University of Auckland with Dr Brown.