Students’ conceptions of assessment: Links to outcomes
Gavin T.L. Brown and Gerrit H.F. Hirschfeld

Students conceive of assessment in at least four major ways (i.e., assessment makes students accountable; assessment is irrelevant because it is bad or unfair; assessment improves the quality of learning; and assessment is enjoyable). A study in New Zealand of 3469 secondary school students’ conceptions of assessment used a self-report inventory and scores from a standardised curriculum-based assessment of reading comprehension. Four inter-correlated conceptions based on 11 items were found with good psychometric properties. A path-model linking the four correlated conceptions with student achievement in reading, while taking into account student ethnicity, student sex, and student year, had good psychometric properties. The conception that assessment makes students accountable loaded positively on achievement while the three other conceptions (i.e., assessment makes schools accountable, assessment is enjoyable, and assessment is ignored) had negative loadings on achievement. These findings are consistent with self-regulation and formative assessment theories, such that students who conceive of assessment as a means of taking responsibility for their learning (i.e., assessment makes me accountable) will demonstrate increased educational outcomes.

Introduction
Conceptions are mental representations of phenomena in reality (Kelly 1991; Thompson 1992), which explain complex and difficult categories of experience, such as assessment (White 1994). Students’ conceptions of educational processes are important because there is evidence that those conceptions have an impact on their educational experiences and learning. For instance, Duit and Treagust (1995) argued that students’ ability to understand science instruction is shaped and limited by their pre-instructional constructions or interpretations of nature. In higher education, students’ learning is more influenced by their perceptions of the educational environment than by the actual educational practices (Entwistle 1991). Furthermore, students’ conceptions of assessment are of particular importance because assessment has a significant impact on the quality of learning (Entwistle and Entwistle 1991; Marton and Säljö 1997; Ramsden 1997).

Pajares (1992) has argued that teachers’ conceptions are a product of their educational experiences as students, suggesting strongly that similar conceptions might be found in both teachers and students. Duit and Treagust (1995) have reported studies wherein teachers passed on their misconceptions of science to students, further suggesting the similarity of conceptions between students and teachers. Thus, the research reported in this paper about students’ conceptions of assessment has grown out of earlier research into teachers’ conceptions of assessment (Brown 2004a) on the assumption that precursor beliefs for teachers’ conceptions might be found among the conceptions of students.

This paper reviews the literature on students’ espoused conceptions of assessment and then reports a study into New Zealand secondary school students’ conceptions of assessment. This

*Corresponding author. Email: gt.brown@auckland.ac.nz
research advances our understanding of students’ conceptions of assessment by linking students’ conceptions with the results of curriculum-aligned assessments of reading comprehension. Furthermore, this research may assist us in understanding the origins of teachers’ conceptions of assessment by showing how students’ conceptions relate to academic outcomes.

**Students’ conceptions of assessment**

Assessment is any act of interpreting information about student performance, collected through any of a multitude of means. Research into the conceptions teachers have about the purposes of assessment has identified four major purposes: that is, (a) assessment improves teaching and learning, (b) assessment makes students accountable for learning, (c) assessment makes schools and teachers accountable, and (d) assessment is irrelevant to education (Heaton 1975; Webb 1992; Torrance and Pryor 1998; Warren and Nisbet 1999; Brown 2002). The research literature on students’ conceptions of assessment is not vast, and is largely focused on tertiary or higher education students (see Struyven et al. 2005 for a review). Review of the empirical literature on students’ conceptions of the purposes of assessment has identified four major purposes, some of which can be matched to teachers’ conceptions of assessment. Students are reported as conceiving of assessment as (a) improving achievement, (b) a means for making them accountable, (c) being irrelevant, and (d) being enjoyable.

**A. Assessment improves achievement and learning**

Pajares and Graham (1998) surveyed 216 Grade 8 students in language arts classes and found that they wanted honest, comprehensible, and constructive feedback on how to improve, while their teachers emphasised praise and positive affect as the important response. In other words, the students wanted to improve and conceived that feedback in response to an assessment should help them to do this. In contrast, their teachers wanted the students to feel good and so denied the students’ access to constructive feedback in order to protect the students from negative consequences. The students were able to see such ‘impression management’ for what it was and sought truth and instruction instead. 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 and Moore 1984). Tertiary students, likewise, have requested more improvement-oriented feedback on their assessed performance than they were actually receiving (Duffield and Spencer 2002).

Perhaps the quality of improvement-oriented feedback is a partial explanation for Davinroy et al.’s (1994) reporting that third-grade students were unable to comment on how their teachers assessed what the students knew about mathematics or reading. Students may not be able to use assessment to improve because teachers do not return work to students for discussion of merits or improvement strategies. In other words, it would appear that students want assessments to contribute to the improvement of educational attainments, but may be frustrated by teachers’ practices and concerns.

**B. Assessment makes students accountable**

Students are frequently assessed in schools and it is quite predictable that a dominant conception of assessment should be making students accountable. Zeidner (1992) offered Israeli junior and senior high school students four purposes for assessment (i.e., summarising student achievement, arousing student interest and motivation, evaluating quality of teaching, and administrative purposes). The students had a stronger perception of assessment as summarising student achievement than as improving motivation or behaviour. Brookhart and Bronowicz (2003: 240) concluded
that, due to the consequences attached to classroom assessments, the 161 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. Even when faced with assessments designed to monitor school effectiveness — for example, high-stakes externally administered assessments (SATs in England), intended to ensure schooling quality — students were found to treat low grades or failure as a measure of themselves rather than of the school or the teacher (Reay and Wiliam 1999).

In the tertiary sector, medical students in Duffield and Spencer’s (2002) research agreed that ideally assessment should meet both school and student accountability (i.e., ensure competence, provide feedback, guide learning, and evaluate curriculum). Maclellan (2001) reported that the majority of both students and lecturers at her university experienced assessment to grade or rank students ‘most frequently’. Indeed, she found that there were statistically significant differences in staff and student perceptions of the frequency of improvement (i.e., assessment to motivate learning, to diagnose) and school accountability (i.e., evaluate teaching) uses; with the students indicating much less improvement use than staff.

Thus, although some students might be aware that assessment could improve or make schools accountable, the conception that assessment makes students accountable is well-entrenched in their thinking.

C. Assessment is irrelevant

Assessment may be considered irrelevant to students if it is thought of as being bad or unfair. Perceived subjectivity and lack of professionalism in the scoring of student assignments or examinations has led tertiary students to view assessment as arbitrary, irrelevant, inaccurate, and simply a necessary process for accruing marks (Duffield and Spencer 2002; Hawe 2002; Sambell et al. 1997). Students are most likely to consider irrelevant assessment practices such as group projects that do not give credit to individual effort (Hoffman and Rogelberg 2001), those that lack explicit criteria for assessment (Maclellan 2001), and end-of-year examinations (Kniveton 1996).

The rejection of subjective assessment has appeared among school students. The attitudes towards assessment of 54 Australian students in their first year of high school became increasingly negative, not only because of the increased volume of assessment compared to primary school, but specifically because of the perceived subjectivity of teacher assessment decisions (Moni et al. 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) to their life chances (Walpole et al. 2005), though in contrast to the English 11-year-olds described above, the tests, rather than themselves, were held responsible for poor results. In fact, partially in response to the high-stakes of accountability assessment, students have treated as irrelevant tasks that they perceive as making high marks more difficult to obtain. Problems that require discovering how to solve them have sometimes been seen by students as awkward, even though students noted that the tasks were intended to require deeper learning (Baird and Northfield 1992; Bell et al. 1992; Shannon and Zawojewski 1995).

Thus, we conclude that students are quite sensitive to assessment which they perceive to be unfair, bad, or irrelevant to them, regardless of their place in the system (i.e., elementary, secondary, and tertiary).

D. Assessment is enjoyable

Much of the literature focuses on students’ attitudes towards different assessment formats or types. For example, Atkinson (2003) reported that Scottish school children enjoyed all the types
of assessment by which they were assessed (including, informal, formal, self, peer, grades, etc.) regardless of purpose (i.e., accountability or improvement). Research into students’ preferences for alternative assessments (such as 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 (Andresen et al. 1994; Slater 1996; Sambell et al. 1997; Segers and Dochy 2001; Bloxham and West 2004). Slater (1996) found that some students thought that they would remember what they learned for longer because the work to produce a portfolio helped them internalise the material.

In multiple studies of high school student evaluations of the Assessment Tools for Teaching and Learning standardisation tests, Hattie et al. (2006) surveyed student opinions about the tests they had just taken. A factor analysis of the responses identified three factors: (1) attitude towards doing the tests, (2) the layout and use of white space, and (3) confidence in doing well on the tests. However, no meaningful correlations between their attitude to these three assessment related factors and their achievement in each subject (i.e., reading, mathematics, pānui [Māori reading], and tuhitahi [Māori writing]) in which they were surveyed were found (mean correlation over ten evaluation factors and four subjects was \( r = .013; SD = .11; \) all statistically significant). Consistent with this, Sambell and McDowell (1998) described, in three case studies, how the same authentic assessments were perceived in opposite manners by students – that is, some students responded with the conception that the assessment improved the quality of learning, while others responded to the same assessment as a means of making students accountable or as something unfair.

Zeidner (1992) found that the 117 teachers and 402 secondary school students he surveyed had very similar perceptions of classroom assessment practices. Both agreed that the goal of grading was to evaluate student achievement and that this should be done through objective testing. Other evidence exists that suggests that students prefer the system of assessment that they experience, regardless of the merits or deficiencies of that system (Blaikie et al. 2004; Deutsch 1949). So the utility of obtaining student opinions about the fun aspect of assessment may not contribute much to our understanding; after all, they may like whatever is in their experience. Thus, the format of assessment may itself be irrelevant to student achievement – it may be students’ conception of the purpose or nature of assessment that is related to different degrees of achievement. Research into the meaning of assessment is fun is taken up in this paper.

New Zealand study of students’ conceptions of assessment

Taken together, the literature identified four major conceptions of assessment from the students’ perspective: that is, assessment acts to improve the quality of learning, assessment is used to make students accountable, assessment is irrelevant, and assessment is enjoyable. At least three of those conceptions were similar to those held by teachers (Brown 2004a). However, there was little evidence available of how students’ conceptions of assessment might interact with academic performance, a matter investigated in this study.

Method

This paper reports the results of a large-scale study into students’ conceptions of assessment. It examines the strength of students’ agreement with different purposes of assessment based on their own experiences of assessment in their lives and links their conceptions of assessment to achievement outcomes on standardised national assessments of literacy. Thus, this research permitted investigation into the nature of the relationships, if any, between students’ conceptions of assessment and their actual curriculum-based performance.
The research programme used self-report survey questionnaire responses and exploratory and confirmatory factor analyses to iteratively refine a set of items that efficiently and accurately identified 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 conceptions results are related to reading achievement outcomes. We conclude with implications for practice and further research.

**Instruments**

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

**Students’ conceptions of assessment – SCoA**

The instrument for eliciting students’ self-reported conceptions of the nature and purpose of assessment was developed from Brown’s (2002) conceptions of assessment instrument used with primary school teachers in New Zealand and Queensland (Brown 2006). The purposes for assessment discovered in that research were used as a basis for determining whether secondary students had the same four main purposes for assessment (i.e., assessment improves education, assessment makes students accountable, assessment makes schools accountable, and assessment is irrelevant). It is worth repeating here that the logic for this procedure rests on the strong likelihood that teachers’ conceptions of assessment were learned as students in school (Pajares 1992) and the possibility that teachers’ conceptions influence students’ conceptions. However, neither of these hypotheses is tested in these studies, as the goal of this research was to explore the students’ conceptions and relate them to student achievement outcomes.

A total of 29 items were selected from a pool of 50 items previously trialled within the context of mathematics achievement (Brown and Hirschfeld 2007). The items were presented in two forms (Form A =19 items; Form B = 21 items) to reduce response fatigue. Eleven items representing four factors with high loadings to mathematics achievement (Brown and Hirschfeld 2007) were common to both forms; three related to ‘assessment makes students accountable’, three to ‘assessment makes schools accountable’, two to ‘assessment is fun’, and three to ‘assessment is ignored’. The other items, not reported here, related to these same factors and to the factor ‘assessment improves education’. However, their factor structure, due to the planned missing data design, was unstable so we have focused on the 11 common items in this paper. A significant threat to the validity of this research is that the 11 items may be affected by their presentation with the assessment of reading and may only provide a partial picture of the full range of conceptions students have. Future research studies will be conducted with a larger pool of items systematically selected to cover a wider range of conceptions and collected independently of the academic performance measure.

The two forms were placed at the end of various trial as Assessment Tools for Teaching and Learning (TTle) reading comprehension tests in the 2004 calibration. The tests containing the conceptions items were administered in 58 schools by the students’ English teachers who were asked to remind students to complete the conceptions items when they had finished the 60-minute test. Students and their parents were informed in writing before the test administration that students had the right not to participate in the asTTle assessment and the following questionnaire study as it was a voluntary research exercise and that any data obtained would be reported anonymously.
Learning outcomes

The outcome measure was secondary school students’ performance on the asTTle 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 et al. 2003). asTTle reports student achievement in each subject using an item response theory (IRT) calibrated scale score (asTTle Mathematics Scale and asTTle Reading Scale) (Hattie 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 test generates scaled scores for total score, curriculum content scores, and cognitive processing scores.

Since the SCoA forms were assigned to different asTTle test forms which had differing test content, the only comparable outcome measures among the datasets was the asTTle total score. The subject total score (aRs) was obtained through 1PL IRT1 calibration of all items and all respondents (Hattie et al. 2004). Although IRT modelling can include item difficulty, discrimination, and pseudo-chance parameters (Embretson and Reise 2000), the student’s ability in asTTle 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. Note that the asTTle scores used in the analysis were the logit scores ($M = 4.29$, $SD = 1.61$), rather than the standardised linear transformed scores (sample $M = 700$, $SD = 100$) since there were problems with the estimated path coefficients and co-variances attributable to the vastly different scales (Kim and Mueller 1978). Thus, this analysis was able to compare students’ conceptions of assessment to their reading comprehension 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 EM missing values procedure (Dempster et al. 1977). 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). As a result, 1803 completed form A, while 1701 completed form B.

The same six-point, positively-packed agreement response scale was used in all studies, as reported in Brown (2004b). This response format is beneficial when it is expected participants are expected to be positively inclined towards all statements by allowing for more shades of positive attitude than would be the case in a balanced response format (e.g., Likert scale).

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 and 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 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 (CFA) 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 and Bentler 1995), which was well exceeded by the 11 common items analysed.
here. Third, structural models were constructed by including outcome measures (i.e., asTTle reading achievement) and student demographic variables (i.e., student sex, student year level, and student ethnicity) into the model. The model of students’ conceptions of assessment based on 11 items had four intercorrelated factors each of which had weights on achievement. The structural model of how the four conceptions factors related to achievement was independently developed with Form A \((N = 1803)\) and then the Form B \((N = 1701)\) data were used separately to cross-validate the solution. As the solutions were statistically similar chance effects due to sample characteristics could be eliminated. CFA was conducted with AMOS (Arbuckle 2003).

**Results**

A total of 3504 participants in the asTTle standardisation of secondary school reading comprehension assessment provided valid responses to either one of two forms of the second student conceptions of assessment inventory (SCoA-II). After dropping participants who had more than 10% of items missing, the sample analysed here was 3469 – a drop of only 35 (1%) students is taken to indicate that there was little fatigue or attention problem in completing the asTTle test or the conceptions of assessment items.

This sample was 54% female compared to the asTTle population of 49% female. Just over half (55%) of the students were of New Zealand European ethnicity \((n = 1699)\) compared to the asTTle population that were 43%; only 13% were Māori (i.e., aboriginal peoples of New Zealand) compared to 29% in the asTTle population. A further 290 students were Pasifika (i.e., a term for people from the island countries of the South Pacific region like Samoa, Tonga, Fiji, etc.) and 736 were Asian or other ethnicity. Thus, at least regarding sex and New Zealand European ethnicity this sample reflected adequately the distributions seen in the asTTle population, which in itself reflected adequately the New Zealand population. By implication, the sample is much less representative for the minority group ethnicities and interpretations by ethnic subgroup may not be valid. There were 1462 Year 9 students (nominally 13 years old), 967 in Year 10 (14-year-olds), 449 in Year 11 (15-year-olds), and 591 in Year 12 (16-year-olds) – all of these values were sufficiently robust. Approximately two-thirds of the sample were enrolled in co-educational schools, all but 53 were in state schools, 20% were in lowest socio-economic schools, and 47% were in highest socio-economic schools. Thus, this large sample is reasonably generalisable to the population of New Zealand secondary school students, though somewhat skewed by ethnicity and socio-economic status.

**Measurement and structural models**

The measurement model consisted of four factors, that is, assessment makes students accountable, assessment makes schools accountable, assessment is fun, and I ignore assessment. The two independent samples (Form A and Form B) provided very similar psychometric properties: item loadings on factors (Table 1), measurement model inter-factor correlations (Table 2A), and fit statistics for independent models (Table 2B).

The high degree of similarity between Form A and Form B values indicated that the measurement and structural models were not subject to chance effects attributable to sampling. Since there was strong similarity between the two independent samples, data are reported for the combined population. Maximum likelihood confirmatory factor analysis of the combined measurement model had good fit characteristics \((\chi^2 = 343.157; df = 38; \text{RMSEA} = .048; \text{TLI} = .95; \text{CFI} = .96)\). Means and standard deviations for the conceptions factors are given in Table 3. Overall the New Zealand high school students in this study moderately agreed that assessment made them accountable, slightly agreed that assessment made schools accountable, slightly
Table 1. Items and measurement model loadings – SCoA-II Form A and Form B.

<table>
<thead>
<tr>
<th>Conception Item</th>
<th>Loadings</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Form A</td>
<td>Form B</td>
</tr>
<tr>
<td><strong>Student accountability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment is assigning a grade or level to my work</td>
<td>.67</td>
<td>.64</td>
<td></td>
</tr>
<tr>
<td>Assessment is checking off my progress against achievement objectives</td>
<td>.69</td>
<td>.64</td>
<td></td>
</tr>
<tr>
<td>Assessment is comparing my work against set criteria</td>
<td>.59</td>
<td>.60</td>
<td></td>
</tr>
<tr>
<td><strong>School accountability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment keeps schools honest and up-to-scratch</td>
<td>.70</td>
<td>.70</td>
<td></td>
</tr>
<tr>
<td>Assessment measures the worth or quality of schools</td>
<td>.61</td>
<td>.58</td>
<td></td>
</tr>
<tr>
<td>Assessment provides information on how well schools are doing</td>
<td>.68</td>
<td>.64</td>
<td></td>
</tr>
<tr>
<td><strong>Assessment is fun</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment is a positive force for improving social climate in my class</td>
<td>.76</td>
<td>.60</td>
<td></td>
</tr>
<tr>
<td>Assessment is an engaging and enjoyable experience for me</td>
<td>.59</td>
<td>.63</td>
<td></td>
</tr>
<tr>
<td><strong>Assessment is ignored</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I ignore or throw away my assessment results</td>
<td>.66</td>
<td>.65</td>
<td></td>
</tr>
<tr>
<td>I do assessments but make little use of the results</td>
<td>.52</td>
<td>.49</td>
<td></td>
</tr>
<tr>
<td>I ignore assessment information</td>
<td>.72</td>
<td>.77</td>
<td></td>
</tr>
</tbody>
</table>

Table 2a. Factor inter-correlations Form A and Form B.

<table>
<thead>
<tr>
<th>Factor</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student accountability</td>
<td>–</td>
<td>.74</td>
<td>.56</td>
<td>−.30</td>
</tr>
<tr>
<td>School accountability</td>
<td>.80</td>
<td>–</td>
<td>.86</td>
<td>−.06</td>
</tr>
<tr>
<td>Assessment is fun</td>
<td>.58</td>
<td>.82</td>
<td>–</td>
<td>−.12</td>
</tr>
<tr>
<td>Assessment is ignored</td>
<td>−.31</td>
<td>−.15</td>
<td>−.13</td>
<td>–</td>
</tr>
</tbody>
</table>

Note. Form A in italics above diagonal; Form B below diagonal.

Table 2b. Fit statistics Form A and Form B.

<table>
<thead>
<tr>
<th>Fit Statistics</th>
<th>Form A</th>
<th>Form B</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1803</td>
<td>1701</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>476.82</td>
<td>423.18</td>
</tr>
<tr>
<td>df</td>
<td>81</td>
<td>81</td>
</tr>
<tr>
<td>CFI</td>
<td>.93</td>
<td>.93</td>
</tr>
<tr>
<td>TLI</td>
<td>.91</td>
<td>.90</td>
</tr>
<tr>
<td>RMSEA</td>
<td>.05</td>
<td>.05</td>
</tr>
</tbody>
</table>

Note. CFI=comparative fit index; TLI=Tucker-Lewis index; RMSEA=root mean square error of approximation.
agreed that assessment was fun, and tended to disagree with the conception that assessment is ignored.

Students’ conceptions of assessment were inter-correlated (Table 4). The school and student accountability factors were correlated at .77. Interestingly, the factor ‘assessment is fun’ correlated very highly with the conception of school accountability ($r = .84$) and moderately with the student accountability factor ($r = .54$). The pattern of these correlations indicated that the students who enjoyed assessment experiences tended to assume that schools rather than students were being made accountable. Perhaps, students’ personal evaluative reaction to assessment forms the basis of their evaluation of school quality; in other words, if students enjoy assessment then their school must be doing a good job. Nevertheless, further detailed investigations with students are required to validate this interpretation. The correlations between the other scales and the ‘assessment is ignored’ conception were quite low, indicating that this factor was a more independent aspect of students’ conceptions.

A structural model containing this measurement model, reading comprehension achievement, and student demographic variables (i.e., student sex, student year level, and student ethnicity) was freely estimated. The structural model values are standardised partial regression weights, such that a value of 1.0 indicates an increase of one standard deviation in the dependent variable. In this model, the four conceptions of assessment along with sex, ethnicity, and school year are simultaneously evaluated for their effect on reading achievement. As some of the respondents had not specified gender or ethnicity, only 3469 cases were analysed. The structural model had good fit characteristics ($\chi^2 = 803.521; df = 81; RMSEA = .051; TLI = .91; CFI = .93$) and all loadings, with $N = 3469$, greater than .05 are statistically significant at alpha = .01 (Figure 1).

The noteworthy features of this structural model were: (a) students’ conceptions of assessment had weak to moderate relations with achievement; (b) only school year exceeded the loading of the four assessment conceptions on achievement; and (c) only one assessment conception positively loaded on achievement, while three were inversely related. Thus, the two biggest and positive predictors of student achievement were school year ($\beta = .52$) and the conception that assessment made students accountable ($\beta = .42$).

Table 4. Combined factor inter-correlations of students’ conceptions of assessment (Form A and B).

<table>
<thead>
<tr>
<th>Factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Student accountability</td>
<td>–</td>
<td>.77</td>
<td>.58</td>
<td>−.30</td>
</tr>
<tr>
<td>2. School accountability</td>
<td>–</td>
<td></td>
<td>.84</td>
<td>−.11</td>
</tr>
<tr>
<td>3. Assessment is fun</td>
<td>–</td>
<td></td>
<td></td>
<td>−.13</td>
</tr>
<tr>
<td>4. Assessment is ignored</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. $N=3504$
In other words, an increase in student year and the conception that assessment was about grading students were positively related with an increase in reading achievement scores. Conversely, male students, non-majority ethnicity students, those who conceived of assessment as fun, those who ignored assessment, or those who conceived of assessment as a measure of school accountability were all associated with lower reading achievement scores. Multiple regression analysis with the same variables indicated that the conceptions accounted for 8.3% of the variance in achievement ($f^2 = .09$ which approaches $f^2 = .15$ for a medium effect; Cohen 1992), on top of the variance explained by sex, year and ethnicity which together explained 30.4% of the variance ($f^2 = .44$ a large effect). Thus, the relationships of students’ conceptions of assessment to educational outcomes were non-chance, and accounted for sufficient variance to be of real interest to educators.

Three conceptions reported here are consistent with three main conceptions identified in the literature; that is, assessment makes students accountable, assessment makes schools or teachers accountable, and assessment is irrelevant. The conceptions were all related, in non-chance ways, to educational achievement. One new conception was identified (assessment is fun), however it is worth noting that due to the use of only 11 items, these results do not capture the full range of students’ conceptions of assessment. However, the two sample design allows us to resist any claims that the structural weights and patterns are chance artefacts due to sampling.

Discussion

Four conceptions of assessment held by students were identified (i.e., assessment makes schools accountable, assessment makes students accountable, assessment is ignored, and assessment is fun). These four conceptions had statistically significant loadings on achievement, accounting for about 8% of outcome variance. The student accountability conception loaded positively on achievement, while school accountability had a similar strength inverse relationship with achievement. The ‘assessment is fun’ and ‘assessment is ignored’ conceptions had negative loadings on achievement and tended to be smaller than those of the two accountability conceptions.

Our approach to interpreting these findings is guided by self-regulation theory – self-regulation of learning concerns a ‘self-directive process through which learners transform their mental abilities into task-related academic skills’ [italics in original] (Zimmerman 2001). Self-regulating students are generally known to achieve more on most educational measures (Ryan et al. 1985; Ryan and Grolnick 1986; Reeve 1998; Reeve et al. 1999); whereas those who locate control or apportion responsibility elsewhere (e.g., assessment makes schools accountable) tend to achieve less on educational outcomes (Rotter 1982). Likewise, students who lack confidence to achieve...
(e.g., assessment is ignored) tend to achieve less (Bandura 1989; Pajares 1996). These results suggested that students who see assessment as a constructive force for personal responsibility gained higher grades, while those who sought to attribute responsibility to schools or teacher for assessment results, those who did not take assessment seriously, or who ignored assessment, received lower grades. Thus, from a naturally occurring population, evidence has been found that students’ conceptions of assessment play a non-chance role in relation to academic performance.

Additionally, the students’ conceptions outlined here and their relationship to achievement, are consistent with formative assessment theories (Crooks 1988; Black and Wiliam 1998; Shepard 2000; National Research Council 2001). Students who conceive of assessment as something that they do not ignore or consider as interfering with their learning are thinking along the lines of formative self-assessment.

Thus, we see in these results students who reported thinking in terms of self-regulation and formative assessment tended actually to achieve more. We further suggest that it is the interaction of both believing in self-responsibility and using assessment formatively that leads to greater educational achievements. The inverse structural weights suggest that one without the other may cancel expected gains. How these conceptions are mediated by actual student practices is an important topic for further studies.

These results may have implications for compulsory, large-scale, testing programmes, such as those envisaged by policies like No Child Left Behind. If the assessment programme is presented to students as a school or teacher accountability mechanism then, according to these data, achievement is likely to go down; whereas, if the assessments are presented as measures of individual student learning, and students believe this, then scores are more likely to go up. It is worth noting that the data were collected in a low-stakes environment – scores on the tests had no impact on either the students or the schools. Thus, further investigation of this finding in a context of high-stakes testing is merited.

Furthermore, these results have implications for teacher preparation. It appears that students who conceive of assessment in terms of personal accountability for learning achieve more. As secondary students change into student teachers and eventually become teachers or instructors, it may be that the onus of responsibility and accountability has to switch from the student to the school or teacher. Brown’s (2004a) research with NZ primary school teachers found that those experienced teachers associated school accountability assessment with assessment for improvement. In the light of self-regulation theory, then it may be that teachers who take responsibility for assessment may be more effective classroom practitioners. Certainly further research is needed to establish the nature of the relationship of teachers’ conceptions of assessment to student achievement, but if taking responsibility in one’s own role is the key to better outcomes, then we will need to develop better mechanisms to help student teachers switch from a view that assessment is about making the learner accountable to the view that assessment makes the instructor accountable. This, in our view, is the real challenge of teachers’ assessment literacy.

A number of further studies are warranted. Further sampling with a larger set of items beyond the 11 common ones reported in this article is needed; such research is already being conducted by the authors. The impact of qualifications systems and student ethnicity on students’ conceptions has not been examined and it may be that such contexts would require a different model of how conceptions of assessment are related to educational outcomes. In-depth qualitative analyses, such as think-aloud protocols, are needed to determine whether the self-regulation interpretation is present in students’ minds as they respond to the items. An investigation across multiple school subjects would also identify whether there is a subject or discipline effect on how assessment is conceived and impacts on achievement. Studies are also needed to determine the relationship of students’ conceptions to their practices. This research has identified a possible chicken-and-egg problem – do students have these conceptions because of their assessment experience or do their
conceptions generate their assessment scores? Investigations into the origins and developmental pathways of achievement and conceptions would richly inform not only this research but also instructional practice.

Nevertheless, this research has shown that robust measurement of students’ conceptions of assessment can be undertaken and that meaningful, non-chance relationships between students’ conceptions and their academic achievement exist. It is clearly suggested by this research that students who maximise their conception of assessment as something that makes them personally accountable, who de-emphasise blaming the school or teacher, who treat assessment seriously, and who pay attention to it will achieve more.

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

Notes on contributors
Gavin Brown, senior lecturer in research methodology, was previously the senior manager of the Assessment Tools for Teaching Learning Project. He is the author of Conceptions of Assessment: Understanding What Assessment Means to Teachers and Students to be published in 2008 by Nova Science.

Gerrit Hirschfeld studied psychology at the University of Münster and is currently pursuing a Ph.D. in cognitive and behavioural neuroscience. He conducted the data analysis for this paper as part of his research internship at The University of Auckland with Dr Brown.

References


