Australian Journal of Educational & Developmental Psychology. Vol 12, 2012, pp. 46-59

Student conceptions of assessment by level of schooling: Further evidence for ecological rationality in belief systems[12]

Gavin Brown
University of Auckland
&
Lois Harris
Central Queensland University

ABSTRACT

Student beliefs about assessment may vary according to the level of schooling. The *Students Conceptions of Assessment* version 6 (SCoA-VI) inventory elicits attitudes towards four beliefs (assessment: improves teaching and learning, measures external factors, has affective impact/benefit, is irrelevant). Using multi-group confirmatory factor analysis, responses of primary (n=100) and high school students (n=134) revealed statistically significant mean score differences. The older students agreed less with improvement, affect/benefit, and external factors conceptions and more with the irrelevance conception. This study provides further evidence that student beliefs about assessment are consistent with how assessment is used in their school environments.

INTRODUCTION

Teacher and student beliefs or conceptualisations of educational processes (e.g., teaching, learning, and assessment) have been shown to significantly predict practices and outcomes across a range of learning domains (Clark & Peterson, 1986; Richardson, 1996; Thompson, 1992). As self-regulation improves student academic outcomes (Boekaerts & Cascallar, 2006; Pintrich & Zusho, 2002; Zimmerman, 2008), it is important to understand beliefs which may facilitate or thwart self-regulation behaviours. To be self-regulated learners within a school context, students often need to utilise feedback obtained from educational assessments. To use such feedback effectively, they must have a personally meaningful purpose for assessment (e.g., it functions as an evaluation of my work), exercise effective skills and strategies in preparing for these assessments, and be able to regulate emotional, cognitive, and behavioural responses once the assessment is finished (Zimmerman, 2008). Therefore, what students think about the nature of assessment, their roles, and the purposes of the

[1] Contact
Associate Professor Gavin T L Brown,
School of Teaching, Learning & Development, Faculty of Education,
The University of Auckland,
Private Bag 92019, Auckland, 1142, New Zealand
Email: gt.brown@auckland.ac.nz

[2] Acknowledgement: The authors would like to acknowledge University of Auckland Faculty Research Development Fund (#3609273/9215) for providing the funding for the Measuring Teachers' Assessment Practices Project and Dr. Jennifer Harnett's support of this research. All data were collected and analysed in accordance with conditions set down by the University of Auckland Human Participants Ethics Committee approvals (HPEC #2008/006, #2008/053). An earlier version of this paper was presented to the Classroom Assessment SIG at the 2011 annual meeting of the American Educational Research Association, New Orleans, LA.

assessment is likely to affect how they respond to and participate in these practices, also influencing their ability and desire to be self-regulated learners.

To better understand the relationship between people's beliefs and behaviours, this study drew on Ajzen's (Ajzen, 1991; Ajzen & Madden, 1986) theoretical framework which suggests that people's beliefs are among the predictors of behaviour. Ajzen's (2005) framework identifies that beliefs people have about a phenomenon, their sense of control or agency within that phenomenon, and their understanding of the social norms towards a phenomenon are key determinants of a person's intentions, purposes, or goals. Consequently, such attitudes are significant contributors to behavioural choices. Hence, human responses to educational processes, such as assessment, depend primarily on their appreciation of what the process is and what it is aimed at achieving.

It is also important to remember that beliefs do not exist within a vacuum; there will always be pressure for beliefs to be rational or coherent within the person's environment (that is, ecological rationality) (Rieskamp & Reimer, 2007). Beliefs about a phenomenon tend to differ according to the environmental constraints imposed on the phenomenon. For example, studies into teacher conceptions of assessment have found different patterns of association between assessment for improved learning and assessment for school accountability in differing regions of the world. Among teachers in Hong Kong and China there was a strong correlation (r >.70) between accountability and improvement (Brown, Kennedy, Fok, Chan, & Yu, 2009; Brown, Hui, Yu, & Kennedy, 2011; Li & Hui, 2007), whereas, the correlations were only weakly positive (r <.30) for New Zealand and Queensland, Australia teachers (Brown, 2004b, 2011; Brown, Lake, & Matters, 2011). These differing correlations likely reflect real world differences between the assessment systems in these regions (that is, high-stakes testing environment versus primarily lower-stakes classroom assessment environment). For example, if you work within a high-stakes testing environment, it may be ecologically rational to adopt the view that this level of accountability does improve student learning in order to remain personally motivated to prepare students for these exams.

Even small differences within New Zealand and Queensland policy contexts between primary and secondary levels of schooling (for example, student achievement is certified in secondary but not primary school) have generated significant differences in mean scores between primary and secondary teachers concerning the nature of assessment (Brown, Lake, & Matters, 2011). Hence, it would be expected that primary and secondary school students themselves would have significantly different understandings of assessment because of the differing consequences attached to it; while doing poorly on assessment in primary school has limited consequences, in secondary school, it may limit one's access to tertiary education and some employment opportunities.

The study reported in this paper shares the results of a New Zealand study that investigated students' conceptions of assessment, comparing and contrasting data from students in upper primary, intermediate, and secondary schooling. It will first review existing research on student beliefs about assessment before explaining the New Zealand context for this study and sharing the study's design, results, and implications.

Student beliefs about assessment

There is a relatively limited body of research relating to student conceptions of assessment. Previous international studies indicate that, while students begin school feeling positive about assessment, negativity increases as they continue their education, demonstrating that assessment can inspire strong affective responses in students. Moni, van Kraayenoord, & Baker (2002) found that Australian students in their first year of high school became increasingly negative towards literacy assessments and suggested that this attitudinal shift was due to increases in the volume and difficulty of assessment, alongside perceptions that teacher assessment decisions were subjective.

Student negativity may also increase because students become more aware of the personal consequences of assessment. Wheelock, Bebell, and Haney's (2000a, 2000b) analysis of grade 4, 8, and 10 students' drawings of their experiences of Massachusetts Comprehensive Assessment System testing found older students were three times more likely to draw negative pictures than younger ones; they hypothesised this was due in part to increased awareness of the personal implications of their test results (i.e., tracking, retention, graduation). Similarly, pictures of assessment drawn by New Zealand high school students were more emotive and negative than those from primary and intermediate pupils (Harris, Harnett, & Brown, 2009). The United Kingdom's Assessment Reform Group (2002)

also found that older students (11 years or older) were more aware of the meaning and consequences of grades, had increased perceptions that teacher grading was unfair. Further, they reported that students older than 11 were more likely to attribute good results to their own ability and effort, while they found younger students tended to attribute performance to external factors (e.g., quality of school/teacher; luck; the teacher likes me; etc.) or practice. As Rotter (1982) has shown, success attribution in external, uncontrollable factors is associated with reduced academic achievement; whereas, endorsement of internally-located, consistent, and controllable causes is associated with increased performance.

However, it is possible that this increased student negativity is not entirely due to the higher consequences of assessment in secondary school. Brookhart and Bronowicz's (2003) study of American primary and high school students suggested that while high school students appeared to cease 'liking' assessment like the primary students, these older students did report appreciating of the importance of these tasks. They hypothesised that 'liking' may morph into an appreciation of the importance of a task; while students may no longer consider assessment to be enjoyable, it may still be seen as a valued activity. Thus, it would appear that, by high school, students have a complex set of beliefs concerning assessment: their emotional stance seems relatively negative, perhaps because of its frequency and importance for life chances, their attributions for success tend to be internally located, and yet assessment is important to them, perhaps because of the opportunity to improve their achievement and proficiency embedded in the feedback system (Butler & Winne, 1995; Hattie & Timperley, 2007).

Together, these studies appear to identify a trend; while students appear initially to accept and perhaps enjoy assessment, student attitudes become more negative as they progress through school, become more aware of the consequences of assessment, and experience higher-stakes assessment practices. In addition to affective response towards assessment, students also have at least four major beliefs about the purposes and nature of assessment (see reviews in Brown & Hirschfeld, 2008; Harris, Harnett, & Brown, 2009). These beliefs included: (1) 'improvement': assessment improves the teacher's teaching and the student's learning; (2) 'external attribution': assessment serves external purposes such as judging a school's quality or predicting students' future educational and employment success; (3) 'affective benefit': assessment is enjoyable and helps classmates be more supportive of each other; and (4) 'irrelevance': assessment is irrelevant because it is unfair and ignored. Research using the *Student Conceptions of Assessment* (SCoA) inventory in New Zealand has shown that the 'improvement' conception has an adaptive effect on standardised test performance in mathematics (Brown & Hirschfeld, 2007; Brown, Peterson, & Irving, 2009) and reading (Brown & Hirschfeld, 2008), and is a belief system aligned with self-regulation theories (e.g., Boekaerts & Corno, 2005; Zimmerman, 2008). In other words, increased endorsement of the 'improvement' conception predicts increased academic performance. In contrast, the more students endorsed the conceptions of 'external attribution' and 'irrelevance' the lower their achievement scores were in both studies. The 'affective benefit' factor (that is, a form of personal well-being emphasis) was inversely related to performance on a standardised mathematics test (Brown, Peterson, & Irving, 2009). The proportion of variance attributed to these factors on one-off tests of academic performance has been moderate (SMC ranged from .13 to .25), indicating that student beliefs about the nature of assessment do matter for individual student academic achievement.

Assessment in New Zealand

To understand the results of the study reported in this paper, it is important to know about the assessment context within New Zealand. Large structural changes have been initiated in New Zealand schooling and education over most of the last three decades (Fiske & Ladd, 2000; Levin, 2001). The national assessment policy in the compulsory school sector (ages 6 to 16) emphasises voluntary, school-based assessment for the purposes of raising achievement and improving the quality of teaching programmes (Ministry of Education, 1994, 2007) relative to the outcomes specified in the national curriculum. The curriculum is child centred, non-prescriptive, holistic, and integrated, but does have specified outcomes and objectives of progression organised into eight levels of performance covering Years 1 to 13 (ages 5-18). There is no compulsory, state mandated assessment regime in the primary school sector (i.e., Years 1-8). All assessment practices in the first ten years of schooling (that is, primary, intermediate, and the first two years of high school) are voluntary and low

stakes. Primary school teachers make extensive use of informal assessments and standardised tests (Crooks, 2002), primarily for the purpose of improving instruction and student learning (Croft, Strafford, & Mapa, 2000; Hill, 2000).

New Zealand subscribes to an Assessment *for* Learning approach (Ministry of Education, 2010), even among secondary schools. Given this approach, students could be expected to be participating in a range of student led assessment practices (e.g., peer and self-assessment), along with standardised measures. Hence, the general culture of school assessment is largely focused on improved teaching and learning, with classroom assessment playing a major role.

However, the national assessment policy has a strong public accountability element. Even primary schools are expected to demonstrate (to parents) that student performance is improving relative to the curriculum levels and objectives. Secondary schooling, however, has a high-stakes, student qualifications system (that is,, the National Certificate of Educational Achievement, NCEA) which begins formally in the third year of secondary schooling when students are about age 15 (Crooks, 2010). The system is composed of both internally and externally assessed items, including tasks which are not 'tests' per se (for example, projects, reports) which are marked using standardised criteria. This means that teachers are actively involved in administering, evaluating, and judging student performance through internal assessments and must also prepare students for end-of-year external examinations. Furthermore, there seems to be a wash-back effect of the national qualification system upon how assessment practices are implemented in the early years of high school. Assessment practices in the first two years of high school (Grades 9 and 10) are officially supposed to be school-based measures used to help students improve learning and report on their progress. However, there is strong evidence that teachers are using NCEA grading systems in those grades as preparation for the official NCEA system (Bashford, 2007; Mizutani, 2006; Rae, 2007).

Hence, one would expect that the impact of assessment on teacher and student beliefs would become more focused on the accountability and external purposes as students transition from a primary to high school. Indeed, New Zealand secondary teachers agreed the conception that assessment makes students accountable much more than primary teachers (Brown, 2011). Likewise, among Year 12 high school students, after their first experience with the qualifications system in Year 11, the relationship between endorsement of student accountability as a purpose for assessment and performance on a standardised test was much stronger than it was for younger students who had not yet experienced the qualifications system (Hirschfeld & Brown, 2009). Thus, it seems likely, then, that students in New Zealand would be committed to the notion of assessment for improved learning and teaching in the primary schooling years, but in secondary schooling students may endorse more strongly the conception that assessment serves certification.

METHOD

The study reported in this paper aimed to expand the usage of the Student Conceptions of Assessment (SCOA) inventory to include younger New Zealand students (primary and intermediate school students) using a non-experimental survey design. The following hypotheses were proposed:

1. It was expected that the measurement model of the SCoA would be sufficient to explain the responses of the primary and intermediate aged students.
2. Because of the essentially formative nature of assessment in New Zealand, it was deemed likely that students across year levels would endorse most strongly the conception that assessment serves to improve student learning.
3. It was hypothesised that younger students would have higher mean scores for improvement than high school students as a consequence of the primary school assessment environment.
4. Although all assessment in New Zealand is intended to be low-stakes until Years 11-13, it was expected, because of wash-back effects, that high school students in Years 9 and 10 would have a stronger endorsement of student conceptions of assessment as a means of accountability than primary school students.

Table 1. Demographic characteristics of students according to school level.

	School Level		
Demographic	Primary	Intermediate	High School
Sex			
Female	17	18	64
Male	32	33	70
Ethnicity			
New Zealand European	42	23	73
Māori	3	2	46
Pasifika (Pacific Ocean island nations)	1	9	4
Asian	2	17	11
Grade			
5	10	—	—
6	29	—	—
7	9	25	—
8	—	26	—
9	—	—	47
10	—	—	87
Mean Age (*SD*)	10 .42 (.66)	11.94 (.84)	14.27 (.65)
School Socio-Economic Status			
High (Deciles 8-10)	27	22	71
Medium (Deciles 4-7)	22	29	23
Low (Deciles 1-3)	0	0	40
Total	49	51	134

Participants

Students were recruited through teachers who participated in Studies 1 and 2 of the *Measuring Teachers' Assessment Practices* (MTAP) project. This project looked at teacher and student conceptions and practices of assessment, focusing on Years 5-10. In all, nine teachers who were already participating in the study agreed to approach students in one or more of their class groups and invite them to participate. To participate, students had to obtain signed parental consent and provide their own written consent in accordance with The University of Auckland Human Participants Ethics Committee approval #2008/053. In all, 347 questionnaires were sent out to teachers for distribution and 234 were returned, comprising a 67% return rate (primary *n*=49, intermediate *n*=51, secondary *n*=134). Responses came from students in Years 6-10. There was some overlap between the ages of the primary and intermediate students as two of the sampled primary classes were composite Year 6/7 classes from full primary schools (that is, schools which cater for students in Years 1-8), while intermediate student questionnaires were from students in Years 7 and 8.

Demographic details of the sample (Table 1) show considerable differences by school level. The proportion of males to females was much higher in the primary and intermediate groups (approximately 2:1), whereas it was nearly equal in the high school group. The majority of students were New Zealand European ethnicity, which accords with New Zealand census statistics. However, the proportion of Māori students across the levels was quite distorted. There are between 15-20% Māori in the New Zealand population, but the indigenous group were underrepresented in the two lower levels and significantly over-represented in the high school group. This latter result was a consequence of one participating school having a high percentage of Māori students. The grade distribution followed the normal expected pattern of Primary (Grades 1-6), Intermediate (Grades 7-8), and High School (Grades 9-13). Only nine students in the Primary sample were in Grade 7. This indicates that inferences that can be made about school level fundamentally reflect the appropriate

grade distributions. Consistent with this distribution was the mean age of students in each school level, with mean scores ranging from about 10½ to 14.

New Zealand schools are classified by the Ministry of Education according to the socio-economic resources of families whose children attend each school. A ten-point scale (deciles) is used, with each decile containing $1/10^{th}$ of the population. It is common to aggregate deciles 1-3 as low socio-economic status (SES), deciles 4-7 as medium, and deciles 8-10 as high. The most notable feature of the distribution is the absence of low socio-economic schools in the Primary and Intermediate groups, whereas the High School group more closely reflected the New Zealand population with 29% in the low SES categorisation.

Instrument

Student beliefs were elicited from self-reported responses to the *Students' Conceptions of Assessment* (SCoA) inventory (Brown, 2003). The questionnaire has 33 items related to four major conceptions of assessment (that is, improvement, irrelevance, external, & affect) (items in Appendix A). The sixth version of the SCoA structures responses into four 2^{nd}-order inter-correlated meta-factors based on a mixture of 1^{st}-order factors and items (Brown, Irving, Peterson, & Hirschfeld, 2009; Brown, Peterson, & Irving, 2009). Students responded using a six-point positively packed agreement rating scale (that is, two negative and four positive responses). Use of positively packed rating scales has been shown to be effective in eliciting variance in responses in social conditions when respondents are likely to be positive about all constructs (Brown, 2004a; Klockars & Yamagishi, 1988; Lam & Klockars, 1982).

Since the SCoA had been designed and used with secondary and tertiary students, small scale studies were carried out to evaluate the accessibility of the inventory for use with younger students. By organising the statements into four prose paragraphs related to the four main factors, it was possible to generate readability statistics for the inventory. Just 9% of the sentences were in passive voice. The Flesch Reading Ease was 61.6 and the Flesch-Kincaid Grade Level was 6.9, suggesting that the inventory should be readily understood by students aged 11 or in Year 6.

Analysis

Before any analyses were carried out, missing data were resolved. There were minimal amounts of missing data (that is, only 43 missing cells across 33 items and 234 students) which were imputed using the expectation maximisation procedure (Dempster, Laird, & Rubin, 1977). Little's MCAR test had χ^2/df=1.27, p=.26, indicating that the missing values analysis produced a data set that was not statistically different to starting values for item means, standard deviations, and covariance.

This study used multi-group invariance testing (Byrne, 1989) to determine whether the existing model for the SCoA applied equally to the current sample and to the three groups of participants (that is, primary, intermediate, and secondary) within the sample. Equivalence of the configuration of pathways (that is, fixed, free, and zero), equivalence of the regression weights and intercepts of items at the latent trait factors, and equivalence of inter-factor covariance matrix were sought before factor or scale means could be compared (Vandenberg & Lance, 2000). Configural equivalence was accepted if the RMSEA index was <.05 and equivalence of regression weights, intercepts, and covariance matrices were accepted if the difference in the CFI index was <.01 (Cheung & Rensvold, 2002). Equivalence was examined only if the model was admissible for all groups (that is, no negative error variances and covariance matrix is positive definite) (Gerbing & Anderson, 1987).

Alternative structures were examined to resolve model inadmissibility, which would be most easily attributed to the small sample sizes per group used here, as samples less than 400 are known to be less robust in their estimation (Boomsma & Hoogland, 2001). The solution for non-positive definite covariance matrices among inter-correlated factors (which is the structure of the SCoA-VI) was to introduce a hierarchical super-order factor structure (that is, Conceptions of Assessment) on the model, thus retaining the original conceptual meaning of factors and item paths. This revision retained the original meaning of the model while generating an admissible and well-fitting solution. In accordance with recommendations for small negative error variances, these were corrected to a small value above zero (that is, .005) (Chen, Bollen, Paxton, Curran, & Kirby, 2001).

Once an admissible model was identified and found to adequately fit the data, the equivalence of parameters was evaluated with multi-group invariance testing. All model analyses were conducted

with AMOS (Arbuckle, 2008) using Pearson product moment correlations; acceptable fit was taken when χ^2/df had $p>.05$, gamma hat >.90, RMSEA and SRMR <.08 (Fan & Sivo, 2007; Marsh, Hau, & Wen, 2004). After demonstrating that the model was invariant across groups within the sample, factor mean scores for the contributing groups were contrasted using multiple analysis of variance.

RESULTS

The data were fitted to the SCoA-VI model with acceptable fit for three of the fit indices (χ^2=954.19, *df*=482, χ^2/df=1.98, *p*=.16; CFI=.87; gamma hat=.89; RMSEA=.065; SRMR=.069). Multi-group confirmatory factor analysis generated a non-positive definite covariance matrix for the high school students. To address this problem, a hierarchical 3rd-order factor was introduced and the resulting model was admissible, after constraining to .005 three negative error variances for primary, six in intermediate, and four in secondary. The revised model had acceptable fit for the χ^2/df, gamma hat, and RMSEA indices (χ^2=2756.33, *df*=1465, χ^2/df=1.88, *p*=.17; CFI=.67; gamma hat=.93; RMSEA=.062; SRMR=.130). These results gave us sufficient warrant to consider that the four factors of the SCoA inventory could be recovered from students in this sample as they had been in previous studies. Nevertheless, the following results require corroboration with much larger samples of students to ensure that the assumptions concerning the negative error variances are valid.

Multi-group invariance testing showed that the model was statistically equivalent as parameters were progressively constrained: equivalent factor to item regression weights (ΔCFI=.000), equivalent regression weights from higher order factors to lower order factors (ΔCFI=.002), equivalent structural covariances (ΔCFI=.000), and equivalent structural residuals (ΔCFI=.007). Hence, the inventory elicited responses from students in an identical fashion regardless of their level of schooling. Figure 1 shows the structure and path values for the constrained equivalent model.

Mean scores for each of the four conceptions were created by averaging all items contributing to the conception. Evaluation of differences in means for the four conceptions was carried out with multiple analysis of variance using one factor (that is, school level) with three groups. Each factor had statistically different means (Table 2) with small effect sizes (Cohen, 1992).

Table 2. MANOVA results for four conceptions of assessment and three levels of schooling

SCoA	*df*	*F*	*p*	R^2	f^2
External	2	14.486	.000	.08	.09
Improvement	2	28.714	.000	.13	.15
Irrelevance	2	16.015	.000	.11	.12
Affect	2	22.024	.000	.11	.12

Tukey HSD post-hoc analysis of observed means showed that the secondary group consistently was a different subset from the primary and intermediate groups (see Table 3). Secondary students gave lower mean scores for ‘external’, ‘improvement’, and ‘affect’, and higher scores for ‘irrelevance’. There were no statistically significant differences between primary and intermediate students, suggesting a major impact on student beliefs may be associated with the transition to secondary schooling. It should be noted that the factor means for the high school students were very similar (absolute value of Cohen’s *d* ranged from .05 to .28 indicating trivial to small differences) to the means reported from an earlier nationally representative sample of New Zealand high school students (Brown, Irving, Peterson, & Hirschfeld, 2009). Hence, insofar as factor means go, it could be argued that this sample of high school students reflects the beliefs of the population of New Zealand high school students and that these high school students endorsed the four conceptions of assessment in a manner quite different from their younger peers.

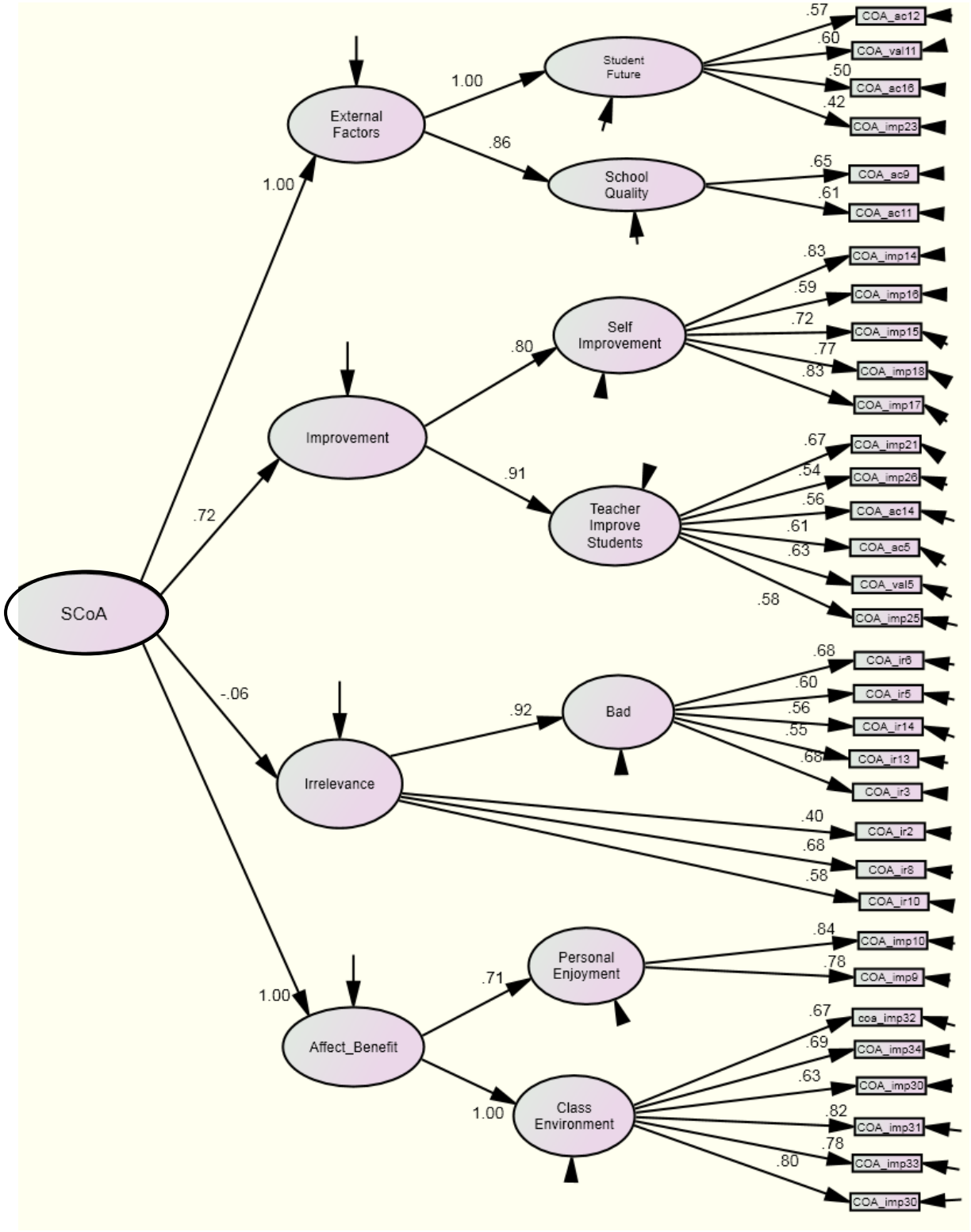

Figure 1. Student Conceptions of Assessment structure with constrained equivalent parameters across three groups. Items are listed in Appendix A.

Table 3. Tukey HSD classification of SCoA observed means by level of schooling

		Student Conceptions of Assessment							
		External Subset		Improve Subset		Irrelevance Subset		Affect Subset	
Level of Schooling	*N*	1	2	1	2	1	2	1	2
3 Secondary	131	3.81		4.14		2.81		2.93	
1 Primary	44		4.34		4.77		2.15		3.71
2 Intermediate	51		4.55		5.05		2.21		3.99
P		1.00	.40	1.00	.11	1.00	.92	1.00	.22

DISCUSSION

This study was the first using the SCoA inventory with children below the secondary level. Results indicated that the inventory elicited responses in a similar fashion to high school age students. This supported our first hypothesis that the measurement model of the SCoA would have sufficient fit to data so as to explain the responses of the upper primary age students. Further, perhaps because of the essentially formative nature of educational assessment in New Zealand and, perhaps because of self-regulatory processes, all students, regardless of grade, endorsed most strongly the conception that assessment serves to improve student learning, as per our second hypothesis. The results appear to be consistent with self-regulation theory. Consistent with a self-regulating approach to assessment, all students endorsed most strongly the 'improvement' factor and gave less than slight agreement to the 'irrelevance' factor. These mean scores suggested that students considered assessment relevant to improved learning, which is an essential component of most self-regulation models (e.g., Zimmerman, 2008).

Nonetheless, the experience of being a high school student had a significant impact on the conceptions students had of assessment consistent with the studies reviewed above. Clearly, the older students, like previous studies with New Zealand students in Grades 9 and 10, had a less 'formative' or improvement-oriented conception of assessment as they agreed less with the improvement and more with the irrelevance conception. Older students also recorded lower agreement with conceptions that assessment provides affective benefit, and is external, suggesting less dependence on external locus of control in their evaluation of the purpose of assessment. In contrast, the younger students endorsed more strongly the formative improvement conception, the positive affective element (enjoyment and improved class morale), and the sense that assessment served external purposes, such as reporting to parents, predicting future employment, showing intelligence, and evaluating their schools. Together, these differences hint that a higher level of self-regulation was apparent among the secondary school students because they appeared to be less dependent on the emotional well-being effects of assessment and more self-reliance in responding to assessment. Thus, there was evidence to support our third and fourth hypotheses that students would respond differently at differing levels of schooling; these differences suggest older students are more self-regulating in relation to assessment. This is not to suggest that primary students do not attempt to be self-regulating in their assessment experiences; it is just that their approach seems more naïve and trusting. They seem to expect that, like all teaching and learning activities, they will enjoy assessment (i.e., it won't be negative experience) and that it is very much a predictor of their own futures. High school students appear to have lost that optimism and perhaps become more realistic about what assessment does.

Due to the vagaries of convenience sampling and voluntary participation, results by school level are confounded by differences in sex, ethnicity, and socio-economic status. Given the extremely small sample sizes for these cells, it is not possible to eliminate these confounds in this study. Previous studies with version II of the SCoA (Hirschfeld & Brown, 2009) showed that there were significant differences in how conceptions of assessment related to academic performance according to sex, ethnicity, and even grade in high school. Thus, grade differences (that is, high school vs. below high school) in this study may reflect sampling differences (that is, sex, ethnicity, and SES) rather than real grade differences. This can only be resolved in future studies with more representative sampling.

Additionally, students' differing views within this sample may not be just self-regulating or rational adaptations to the higher-stakes consequences of educational assessment in secondary school. It is feasible that adolescent developmental processes and related factors (for example, egocentrism, increased self-oriented concept of oneself, or identity formation, see Durkin, 1995) are also contributing to the changes in the structure of students' understanding of assessment.

Another threat to the validity of the findings is the instability of the model estimates in the multi-group condition. Sample size is the most likely explanation for this instability because the original model worked with the combined groups without negative error variances or non-positive definite covariance matrix. Ensuring accuracy in estimation can only be resolved in future studies with much larger samples for each demographic characteristic of interest.

Nonetheless, this study is a cross-sectional analysis of different age groups and is confounded by the influence of different teachers, schools, socio-economic strata, and ethnicities on student conceptions of assessment. Longitudinal tracking of student conceptions of assessment across the high school transition is clearly warranted as there is strong evidence that appropriate beliefs about assessment contribute significantly to academic performance. Furthermore, the relationship of the different conceptions to academic performance for younger students needs to be examined to identify potential adaptive and maladaptive effects that may be different from those of older students.

Hence, this analysis showed that level of schooling may be a significant predictor of student conceptions of assessment, with the transition to secondary schooling significantly shaping student beliefs within a New Zealand context. Putting aside the differences in sampling which could be resolved with larger scale studies, it would appear that the implementation of educational assessment in high schools, along with the higher-stakes consequences associated with these practices, may be a contributing factor to adolescents having different beliefs from their younger peers. Studies with New Zealand and Queensland teachers have shown that high school teachers place more emphasis on assessment as student accountability than primary school teachers (Brown, 2011; Brown, Lake, & Matters, 2011). This suggests that the introduction of student qualifications or certification assessments in high school likely has an impact on the beliefs of both teachers and students. Perhaps, the shift in conceptions is a rational response to real changes in how assessment is implemented in high schools. The study provides some evidence for the ecological rationality argument about the nature of belief systems; students' beliefs appear to be consistent with the ecological framework of assessment purposes and uses in the two different levels of schooling.

As the potential for negative consequences increases in high school, it is reasonable to expect students to become more negative about the role assessment plays in their lives. It is likely that this will be more noticeable among lower performing students (Assessment Reform Group, 2002).

Nonetheless, improvement is still the most strongly endorsed conception of assessment among high school students, suggesting that there is still a window of opportunity for teachers. Teachers who actively and explicitly use assessment to improve student learning and their own teaching should be reassured that this is what all students, including high school students, expect from assessment.

REFERENCES

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50,* 179-211.

Ajzen, I. (2005). *Attitudes, personality and behavior* (2nd ed.). New York: Open University Press.

Ajzen, I., & Madden, T. J. (1986). Predictions of goal-direct behavior: Attitudes, intentions, and perceived behavioural control. *Journal of Experimental Social Psychology, 22,* 453-474.

Arbuckle, J. L. (2008). AMOS (Version 17.0.0) [computer program]. Crawfordville, FL: Amos Development Corporation.

Assessment Reform Group. (2002). *Testing, motivation, and learning*. Cambridge, UK: University of Cambridge Faculty of Education.

Bashford, J. (2007). Students' perceptions of the value of different types of teacher feedback. In E. R. Peterson & S. E. Irving (Eds.), *Conceptions of assessment and feedback* (pp. 32-38). Wellington, NZ: Ministry of Education, NZCER Teaching and Learning Research Initiative.

Boekaerts, M., & Cascallar, E. (2006). How far have we moved towards the integration of theory and practice in self regulation? *Educational Psychology Review, 18*(3), 199-210. doi:10.1007/s10648-006-9013-4

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology: An international review, 54*(2), 199-231.

Boomsma, A., & Hoogland, J. J. (2001). The robustness of LISREL modeling revisited. In R. Cudeck, S. Du Toit & D. Sorbom (Eds.), *Structural equation modeling: Present and future* (pp. 139-168). Lincolnwood, IL: Scientific Software International.

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. (2003-2008). *Students' conceptions of assessment (SCoA) inventory* (Versions 1-6). Unpublished test. Auckland, NZ: University of Auckland.

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

Brown, G. T. L. (2004b). Teachers' conceptions of assessment: Implications for policy and professional development. *Assessment in Education: Principles, Policy & Practice, 11*(3), 301-318.

Brown, G. T. L. (2011). Teachers' conceptions of assessment: Comparing primary and secondary teachers in New Zealand. *Assessment Matters, 3*, 45-70.

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

Brown, G. T. L., & Hirschfeld, G. H. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy and Practice, 15*(1), 3-17.

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 & Instruction, 19*(2), 97-111.

Brown, G. T. L., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy and Practice, 16*(3), 347-363.

Brown, G. T. L., Lake, R., & Matters, G. (2011). Queensland teachers' conceptions of assessment: The impact of policy priorities on teacher attitudes. *Teaching and Teacher Education, 27*(1), 210-220. doi:10.1016/j.tate.2010.08.003

Brown, G. T. L., Peterson, E. R., & Irving, S. E. (2009). Self-regulatory beliefs about assessment predict mathematics achievement. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.) *Student perspectives on assessment: What. students can tell us about assessment for learning* (pp. 159-186). Charlotte, NC: Information Age Publishing.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*(3), 245-281.

Byrne, B. M. (1989). Multigroup comparisons and the assumption of equivalent construct validity across groups: Methodological and substantive issues. *Multivariate Behavioral Research, 24*(4), 503-523.

Chen, F., Bollen, K. A., Paxton, P., Curran, P. J., & Kirby, J. B. (2001). Improper solutions in structural equation models: Causes, consequences, and strategies. *Sociological Methods & Research, 29*(4), 468-508.

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

Clark, C., & Peterson, P. (1986). Teachers' thought processes. In M. Wittrock (Ed.), *Handbook of research on teaching.* (3rd ed., pp. 255-296). New York: MacMillan.

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

Crooks, T. J. (2002). Educational assessment in New Zealand schools. *Assessment in Education: Principles, Policy & Practice, 9*(2), 237-253.

Crooks, T. J. (2010). Classroom assessment in policy context (New Zealand). In B. McGraw, P. Peterson & E. L. Baker (Eds.), *The international encyclopedia of education* (3rd ed., pp. 443-448). Oxford, UK: Elsevier.

Croft, A. C., Strafford, E., & Mapa, L. (2000). Stocktake/evaluation of existing diagnostic tools in literacy and numeracy in English. Wellington, NZ: NZCER.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society, Series B, 39*(1), 1-38.

Durkin, K. (1995). *Developmental social psychology: From infancy to old age*. Malden, MA: Blackwell Publishers.

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

Fiske, E. B., & Ladd, H. F. (2000). *When schools compete: A cautionary tale*. Washington, DC: Brookings Institution Press.

Gerbing, D. W., & Anderson, J. C. (1987). Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternate respecifications. *Psychometrika, 52*(1), 99-111.

Harris, L. R., Harnett, J., & Brown, G. T. L. (2009). Assessment from students' perspectives: Using pupil drawings to examine their conceptions of assessment. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.) *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 53-83). Charlotte, NC: Information Age Publishing.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Hill, M. (2000). Dot, slash, cross: How assessment can drive teachers to ticking instead of teaching. *SET: Research information for teachers*(1), 21-25.

Hirschfeld, G. H. F., & Brown, G. T. L. (2009). Students' conceptions of assessment: Factorial and structural invariance of the SCoA across sex, age, and ethnicity. *European Journal of Psychological Assessment, 25*(1), 30-38.

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.

Levin, B. (2001). *Reforming Education: From Origins to Outcomes*. London: RoutledgeFalmer.

Li, W. S., & Hui, S. K. F. (2007). Conceptions of assessment of mainland China college lecturers: A technical paper analyzing the Chinese version of CoA-III. *The Asia-Pacific Education Researcher, 16*(2), 185-198.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*(3), 320-341.

Ministry of Education. (1994). *Assessment: Policy to practice*. Wellington, NZ: Learning Media.

Ministry of Education. (2007). *The New Zealand Curriculum for English-medium teaching and learning in years 1-13*. Wellington, NZ: Learning Media.

Ministry of Education (2010). *Ministry of Education Position Paper: Assessment [Schooling Sector]*. Wellington, NZ: Ministry of Education.

Mizutani, S. (2006, July). *Understanding washback of NCEA on the teaching and learning of Japanese - Not as straightforward as you think?* Paper presented at 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.

Pintrich, P. R., & Zusho, A. (2002). The development of academic self-regulation: The role of cognitive and motivational factors. In A. Wigfield & J. S. Eccles (Eds.), *Development of achievement motivation* (pp. 249-284). San Diego: Academic Press.

Rae, D. (2007). Types of feedback. In E. R. Peterson & S. E. Irving (Eds.), *Conceptions of assessment and feedback* (pp. 73-82). Wellington, NZ: Ministry of Education, NZCER Teaching and Learning Research Initiative.

Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula, T. J. Buttery & E. Guyton (Eds.), *Handbook of research on teacher education* (2nd ed., pp. 102-119). New York: Macmillan.

Rieskamp, J., & Reimer, T. (2007). Ecological rationality. In R. F. Baumeister & K. D. Vohs (Eds.), *Encyclopedia of Social Psychology* (pp. 273-275). Thousand Oaks, CA: Sage.

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.

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.

Wheelock, A., Bebell, D., & Haney, W. (2000a). Student self-portraits as test-takers: Variations, contextual differences, and assumptions about motivation. *Teachers College Record, ID Number: 10635*.

Wheelock, A., Bebell, D., & Haney, W. (2000b). What can student drawings tell us about high-stakes testing in Massachusetts? *Teachers College Record, ID Number: 10634*.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal, 45*(1), 166-183.

Biographical Notes:

Gavin Brown is an Associate Professor in the Faculty of Education at the University of Auckland. His research interests focus on the social psychological impact of assessment on teachers and students. He is a co-editor and contributor to the 2009 book *Student perspectives on assessment*, published by Information Age Publishing.

Lois Harris is an Honorary Research Fellow at the University of Auckland, also lecturing at CQ University Australia and teaching for Education Queensland. Her research interests include assessment, student engagement, and research methodology.

APPENDIX A.
SCoA items

Code	Item
COA_ac11	Assessment provides information on how well schools are doing
COA_ac12	Assessment is important for my future career or job
COA_ac14	Assessment measures show whether I can analyse and think critically about a topic
COA_ac16	Assessment results show how intelligent I am
COA_ac5	Assessment is checking off my progress against achievement objectives or standards
COA_ac9	Assessment measures the worth or quality of schools
COA_imp10	Assessment is appropriate and beneficial for me
COA_imp14	I look at what I got wrong or did poorly on to guide what I should learn next
COA_imp15	I make use of the feedback I get to improve my learning
COA_imp16	I pay attention to my assessment results in order to focus on what I could do better next time
COA_imp17	I use assessments to identify what I need to study next
COA_imp18	I use assessments to take responsibility for my next learning steps
COA_imp21	My teachers use assessment to help me improve
COA_imp23	Assessment tells my parents how much I've learnt
COA_imp25	Assessment helps teachers track my progress
COA_imp26	Teachers use my assessment results to see what they need to teach me next
COA_imp29	Assessment encourages my class to work together and help each other
COA_imp30	Assessment motivates me and my classmates to help each other
COA_imp31	Our class becomes more supportive when we are assessed
COA_imp32	When we do assessments, there is a good atmosphere in our class
COA_imp33	Assessment makes our class cooperate more with each other
COA_imp34	When we are assessed, our class becomes more motivated to learn
COA_imp9	Assessment is an engaging and enjoyable experience for me
COA_ir10	I ignore assessment information
COA_ir13	Teachers are over-assessing
COA_ir14	Assessment results are not very accurate
COA_ir2	Assessment has little impact on my learning
COA_ir3	Assessment interferes with my learning
COA_ir5	Assessment is unfair to students
COA_ir6	Assessment is value-less
COA_ir8	I ignore or throw away my assessment results
COA_val11	Assessment results predict my future performance
COA_val5	Assessment is a way to determine how much I have learned from teaching