

## RESEARCH ARTICLE

# The effect of conceptions of assessment upon reading achievement: An evaluation of the influence of self-efficacy and interest

Gavin T.L. Brown and K. Frank Walton

School of Learning, Development & Professional Practice, Faculty of Education & Social Work, The University of Auckland, New Zealand

## Abstract

Self-regulation of learning requires that students conceive of assessments as a means of reflecting upon and guiding their learning. The relationship of student beliefs about the importance, usefulness, or purpose of assessment to self-efficacy and interest and their joint effect on reading performance has not been investigated. In the context of a large nationally representative survey of New Zealand secondary school students, participants completed either Form 1 or 2 of the Student Conceptions of Assessment (SCoA) inventory version 2, a brief inventory on self-efficacy and interest in reading, and a standardised reading achievement test. Measurement models for both forms of the SCoA were established using exploratory and confirmatory factor analysis. A structural model relating conceptions of assessment to reading performance for each version of the SCoA inventory was established. Invariance of the models for students with high vs. low levels of self-efficacy or interest in reading was tested. Only two conceptions of assessment had statistically significant relations to achievement (i.e., assessment makes me accountable and assessment is useless). Metric equivalence was found for all groups and forms, except version 2A interest. Accountability effects were generally small and not statistically significant, while effects from useless were stronger and negative. Differences between levels of interest and self-efficacy were small. These results suggest that students with lower and higher self-reported interest and self-efficacy can be treated similarly, with a focus on reducing the maladaptive effects of believing assessment is useless.



## Open Access

**Citation:** Brown, G.T., Walton, K. F. (2017). The effect of conceptions of assessment upon reading achievement: An evaluation of the influence of self-efficacy and interest. *Interdisciplinary Education and Psychology*, 1(1):3.

Received: June 03, 2017

Accepted: July 12, 2017

Published: July 24, 2017

**Copyright:** © 2017 Brown et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

## Corresponding author:

Gavin T. L. Brown, School of Learning, Development & Professional Practice, Faculty of Education & Social Work, The University of Auckland, New Zealand.  
E-mail: gt.brown@auckland.ac.nz

## Keywords

Assessment; conceptions; self-efficacy; interest; reading; achievement;

Since doing well on assessment matters to academic outcomes, it is logical that higher-performing students regulate their preparation for, action during, and responses after assessments. Such behaviours are consistent with self-regulation of learning (SRL) theories (Boekaerts, 1995; Boekaerts & Cascallar, 2006; Marsh, Hau, Artelt, Baumert, & Peschar, 2006; Schunk & Ertmer, 2000). Research has shown that student beliefs about assessment contribute in both adaptive and maladaptive ways to performance (Brown, Peterson, & Irving, 2009). However, conceptions of assessment do not act alone in creating the self-regulating learner. Belief that one is good at a subject (i.e., self-efficacy) and interest in the domain being assessed also contribute to greater or lesser outcomes. Hence, this paper examines how greater or lesser self-efficacy and interest interact with conceptions of assessment as predictors of greater or lesser outcomes on a reading test. Such insights will help with the preparation of students depending on whether they have high or low levels of reading self-efficacy and interest.

## Conceptions of Assessment

Students who view assessment as an opportunity to measure their progress against desired learning goals, value the feedback information they gain, and seek to close gaps between their goals and their present performance are regulating their learning (Zimmerman, 2001). Such students submit themselves to the scrutiny of evaluation so that they can receive information about their strengths and weaknesses and guidance as to where to focus their efforts. It is assumed that such students are better able to cope with the stresses of assessment since they accept that, although testing may not be an enjoyable experience, it can assist their learning; what Boekaerts and Corno (2005) refer to as a growth pathway as opposed to an ego-protective pathway. Furthermore, it is presumed that students who do not value the role of assessment as a tool to improve their learning are likely to make excuses for poor performances; for example, blaming their limited intelligence (Blackwell, Trzesniewski, & Dweck, 2007) or more extrinsically the poor quality of teaching (Weiner, 2000). This pattern of attitudes could lead to intentional withholding of effort and reduced outcomes (Boekaerts & Corno, 2005; Hattie, 2004). Hence, it is logical to presume that student beliefs about assessment can function as contributors to greater achievement.

Research into student conceptions of assessment is relatively novel and few studies have identified how those conceptions relate to achievement (McMillan, 2016). Studies with secondary school students in New Zealand have found that some conceptions of assessment are associated with greater and some with lower achievement and that those patterns seem consistent with SRL theory (Brown, 2011). For example, greater academic performance was noted when students agreed more that: (a) assessment made them accountable (Walton, 2009; Brown & Hirschfeld, 2005, 2007, 2008) and (b) assessment improves teaching and learning (Brown, Peterson, & Irving, 2009; 'Otunuku, Brown, & Airini, 2013). In contrast, negative relations to achievement were found when students agreed more that (a) assessment is bad, unfair, or irrelevant (Brown, 2011), (b) assessment is supposed to be fun, enjoyable, or improve classroom climate (Brown, Peterson, & Irving, 2009), and (c) assessment evaluates school quality or predicts students' futures (Brown, Peterson, & Irving, 2009). The positive associations reflect adaptive (i.e., associated with increased learning outcomes) self-regulating responses to assessment, while the negative associations are consistent with maladaptive responses to assessment.

Conceptions of assessment have been found to predict a substantial (i.e., 20-25%) proportion of the variability on standardised tests of academic achievement (Brown & Hirschfeld, 2008; Brown, Peterson, & Irving, 2009). In an American university student study, Wise and Cotten (2009) found less guessing (i.e., longer response times) took place when students agreed more that assessment leads to improvement, while more guessing took place as students agreed more that assessment was irrelevant.

Not surprisingly, the research on student beliefs about assessment has relied upon self-report questionnaires. One of the important properties of a test or inventory is that it provides valid measures of a specified trait, so that differences in achievement or performance between groups reflects real differences in the tested ability rather than construct-irrelevant factors. For example, Hirschfeld and Brown (2009) found measurement invariance for beliefs about assessment according to three demographic attributes (i.e., student sex, ethnicity, and age) and suggested that personal self-motivational beliefs may be of greater interest than demographic variables in understanding factors shaping the impact of student conceptions of assessment. Thus, this paper investigates the degree to which students' control beliefs about assessment are impacted by their self-reported self-efficacy and interest in the subject being tested.

## Self-Efficacy & Interest

Most studies have shown that both self-efficacy and personal interest have a small to medium positive effect on academic performance (Hattie, 2004; Marsh, Hau, Artelt, Baumert, & Peschar, 2006; 'Otunuku & Brown, 2007; Schunk, 1983). An important competence belief (Schunk & Zimmerman, 2006) is self-efficacy (i.e., "conviction that one can successfully execute the behaviour required to produce the outcomes" [Bandura, 1977, p. 79]), which influences the actions people choose and persist with, even in face of difficulties. Importantly, self-efficacy is task and situation specific (e.g., doing a reading test vs. reading a novel) and is normally

generated as a consequence of mastery experiences within specific domains (e.g., reading for meaning) (Bong, 2013). Self-reported levels of self-efficacy in various school subjects have a weak to moderate impact on test scores ( $.20 < \beta < .55$ ). However, these effects differ by school subject and by student overall ability (i.e., self-efficacy tends to be more influential for lower-achieving students) (Bong, 2013).

Another important control belief that influences the processes of SRL is interest in the material being learned or assessed (Zimmerman & Schunk, 2004). The model of domain learning proposes that early in the learning process, when student knowledge or competence is relatively low, situational interest is useful in motivating students to learn (Alexander, 1995). Situational interest refers to establishing in the learning environment a high degree of relevance or situatedness for learning objectives and content in the life, interests, or motivations of the learners. As knowledge competence grows, individual interest develops, sustaining internal motivation to learn an increasing complex and sophisticated understanding of the domain. Hence, students who have a high level of personal interest in a subject area or task tend to display high levels of engagement and enthusiasm, are willing to spend more time on a task, and persist when facing difficult challenges (Hidi & Harackiewicz, 2000).

The aim of this study was to find out how reading self-efficacy and interest interacted with conceptions of assessment and performance on a reading comprehension test. It was expected that the pathways from student conceptions of assessment to academic performance would not be equivalent for students with high and lower levels of self-efficacy and interest. Since conceptions of assessment can have either positive or negative relationship with achievement, we expected that the conceptions with positive association with performance (e.g., assessment improves learning) would show much stronger path values for students with high interest and self-efficacy. In contrast, we expected students with low self-efficacy and interest to have much stronger path values on the maladaptive conceptions of assessment (e.g., assessment is irrelevant, assessment is fun). Thus, we expected an interaction effect between adaptive and maladaptive conceptions of assessment and student interest and self-efficacy, rather than a consistent main effect.

## Method

This study involved secondary analysis of self-report data concerning students' conceptions of assessment collected in conjunction with the national norming of the Assessment Tools for Teaching and Learning (asTTle) reading comprehension test items (Hattie et al., 2004). While the data may be seen as dated, there is little evidence from the international PISA studies that New Zealand secondary student performance in reading has changed since this data being collected (Comparative Education Research Unit, 2016). Furthermore, subsequent studies of New Zealand students have suggested that there have been few shifts in student beliefs about assessment (Brown, 2013; Brown, Peterson, & Irving, 2009; Otunuku, Brown, & Airini, 2013).

### Context

New Zealand has implemented a standards-based national qualifications system as the basis for determining student leaving certificates and entry to higher education (Crooks, 2010). The National Certificate of Educational Achievement (NCEA) system has three levels, with Level 1 introduced in the third year of high school (Year 11), Level 2 in Year 12, and Level 3 in Year 13. Students accumulate credits on a mixture of school-based coursework assignments and end-of-year externally-administered examinations which are aligned to curriculum based standards and objectives. Criterial descriptions exist for standards within each subject and for the levels of achievement (i.e., Not Achieved, Achieved, Merit, and Excellence). School teachers of students in Years 11-13 teach standards content and administer and mark school-based coursework.

Unsurprisingly, teachers introduce Year 9 and 10 students to the NCEA style of grading and actively involve students in peer and self-assessment against the levels of achievement criteria (Harris & Brown, 2013). While intended to be formative, these assessment practices have been shown to be especially helpful to students who are committed to doing their best, rather than doing just enough (Meyer, McClure, Walkey, Weir, & McKenzie, 2009). Hence, it is practically a universal norm that students in New Zealand high schools are exposed to the practice of evaluating their school work against criteria, standards, and levels.

### Participants

Data were obtained from 3803 students (grades 9 – 12) from 58 different secondary schools in

New Zealand (Table 1). The demographic break-down of students by sex, ethnicity, and school grade in the sample approximately represented the proportions in the New Zealand secondary student population. The year levels of participating students was weighted towards Year 9 and 10 students (i.e., 70.5%).

Table 1. Total Student Participant Demographic Breakdown

Demographic category	<i>n</i>	%
Sex		
Boys	1759	46.3
Girls	2044	53.7
Ethnicity		
European	2110	56.5
Māori	523	14.0
Pasifika	316	8.5
Asian	318	8.5
Others	468	12.5
(Not stated)	(68)	(1.8)
Grade level		
Year 9	1622	42.7
Year 10	1059	27.8
Year 11	492	12.9
Year 12	630	16.6

*Note: "Pasifika" = students of Pacific Island ethnicity, predominantly Samoan, Tongan, and Cook Island*

### Instruments

The three instruments used to collect data were the Students' Conceptions of Assessment (SCoA II), the Students' Attitudes to Reading (SAR), and the standardised asTTle reading comprehension tests.

**SCoA II.** The SCoA II instrument consisted of 29 items arranged into two forms to reduce fatigue. The items were adapted from an earlier study (Brown & Hirschfeld, 2007) and results for just 11 common items have already been reported in Brown and Hirschfeld (2008). This paper reports an original analysis of results derived from Walton (2009) of Form 1 (SCoA-IIA) which contained 20 items and Form 2 (SCoA-IIB) which contained 21 items. Table 5 gives the items and indicates which ones are common across forms and the ten items which had been previously reported in Brown & Hirschfeld (2008) (i.e., three in factor assessment makes me accountable, three in assessment makes schools accountable, one in assessment is enjoyable, and three in assessment is useless). Given the similarity of items, it was expected that the two forms would have similar factors and structures.

The questionnaires used a positively packed six-point agreement response scale, with two negative options (strongly disagree, usually disagree) and four positive options (slightly agree, moderately agree, usually agree, strongly agree) (Brown, 2004). Since it was expected that students would rate the various conceptions positively, positive packing was used to increase variance in students' responses and provide more precision in the analysis of the responses (Lam & Klockars, 1982).

**SAR.** Six items elicited motivational attitudes to reading. Self-efficacy in this study refers to perceptions of competence in an academic domain (i.e., completing an objectively scored reading comprehension test about unseen short reading passages). Thus, students are likely to have to infer their self-efficacy for learning from similar prior attainments, because they are unable to know in advance the specifics of the tasks involved (Pajares, 1996). Such prior information would arise from previous test scores or feedback from teachers or parents about the quality of their reading. Thus, in accordance with Pajares' (1996) conclusion, we consider that at

the “self-efficacy for learning levels of generality, self-concept and self-efficacy beliefs may be empirically similar” (p. 563). Hence, the three self-efficacy items were:

- How good do you think you are at reading?;
- How good does your teacher think you are at reading?’ and
- How good does your mum or dad think you are at reading?

**The three interest items were:**

- How much do you like doing reading at school?
- How much do you like doing reading in your own time (not at school)? and
- How do you feel about going to the library to get something to read?

‘Otunuku and Brown (2007) reported, using confirmatory factor analysis, that these items formed two scales (i.e., interest and self-efficacy) with good fit properties. The moderate inter-correlation ( $r=.64$ ) between self-efficacy and interest in reading comprehension indicated each scale could be used separately. Students responded using a four point scale, identified by smiley face symbols indicating degree of affect (i.e., very happy face=4, happy face=3, sad face=2, very sad face=1).

**asTTle Reading.** Academic performance in reading comprehension was determined by performance on norming test forms for the asTTle testing system (Hattie et al., 2004). The items were aligned to the New Zealand national English curriculum levels and objectives (Hattie, Brown, & Keegan, 2003; Ministry of Education, 2007) and scoring was done using single parameter item response theory. This meant that regardless of test form completed, student performance in reading was on a common transformed scale (Embretson & Reise, 2000). The asTTle scores were transformed to a standardised score with a mean for Year 6 set at 500, with a standard deviation of 100 (Hattie et al., 2004).

Table 2 provides the average reading score by year group and SCoA Form for high and low levels of interest and self-efficacy. While there is clearly a strong relationship between student year and reading score, the relationship of being in a high or low level group to reading score, while statistically significant, had small effect sizes (i.e.,  $\eta^2<.13$ ) (Interest  $F(3)=70.053^{***}$ ;  $\eta^2=.07$ ; Self-Efficacy  $F(3)=71.01^{***}$ ;  $\eta^2=.12$ ). Thus, reading score is largely independent of level of interest and self-efficacy, consistent with a previous study (‘Otunuku & Brown, 2007).

Table 2. Reading Score by Form and Level of Interest or Self-Efficacy

Reading Score					Reading Score				
Interest	Year	N	M	SD	Self-Efficacy	Year	N	M	SD
Form 1					Form 1				
High	9	300	691.19	104.99	High	9	226	694.18	100.63
	10	211	773.48	79.32		10	143	782.49	69.16
	11	67	785.55	65.64		11	48	796.05	65.47
	12	96	792.58	56.25		12	83	803.47	52.75
Low	9	289	600.19	81.67	Low	9	219	585.37	77.28
	10	203	739.40	73.98		10	174	731.79	75.90
	11	115	753.68	64.28		11	62	747.32	64.53
	12	123	762.51	56.90		12	90	757.98	58.50
Form 2					Form 2				
High	9	317	675.74	91.75	High	9	238	679.69	90.53
	10	171	771.88	75.88		10	114	781.68	73.02
	11	71	790.85	66.74		11	59	786.95	64.26
	12	123	803.37	62.86		12	107	800.56	66.72
Low	9	266	587.28	77.28	Low	9	198	573.41	77.91
	10	166	699.44	86.79		10	144	680.77	90.43
	11	104	754.10	58.16		11	61	737.59	68.41
	12	120	763.76	58.01		12	91	757.91	59.91

### Data Collection Procedures

For each of the four year levels, multiple as TTle reading tests were prepared, each containing items within an estimated appropriate range of difficulty. At the end of each test, either the Form 1 or Form 2 SCoA questionnaire was attached. It was intended that all test papers would have an equal chance of being assigned to any individual in any class so that any effect of the class or



teacher on the distribution would be randomised. The teachers who supervised the tests were asked to remind students to complete the SCoA questionnaires when they had finished the one hour as TTle test. Student demographic information gathered from the as TTle test included sex, ethnicity, and Year level.

### Data analysis

**Data Preparation.** The Form 1 and Form 2 SCoA data sets were cleaned. First, all participants who had given the same response (e.g., all slightly agree) for 15 or more of all SCoA items were removed, on the assumption that such responding indicated lack of engagement with the substance of the items. This removed  $n=62$  (3.2%) and  $n=88$  (4.8%) of all SCoA Form 1 and Form 2 cases, respectively. Secondly, cases with more than 10% missing responses were removed (Form 1  $n=119$ , 6.3%; Form 2  $n=147$ , 8.3%). After removing between 9-12% of invalid responses, there were still some missing responses, presumably through random inattention.

From the 1774 cases in Form 1, 540 (1.5%) missing responses were observed across the 20 SCoA variables, while in Form 2, there were 492 (1.4%) missing responses from 1623 cases across 21 SCoA variables. Thus, thirdly, these remaining missing values were imputed using the expectation maximisation procedure (Dempster, Laird, & Rubin, 1977). Little's MCAR test was statistically significant for both forms (Form 1:  $\chi^2=1485.06$ ,  $df=1334$ ,  $p=.002$ ; Form 2:  $\chi^2=1850.10$ ,  $df=1655$ ,  $p=.001$ ). However, the  $\chi^2$  test is extremely sensitive with large sample sizes (Tanaka, 1987), which applies in this study. Consequently, the ratio of  $\chi^2$  to  $df$  was examined and found to be not statistically significant (Form 1:  $\chi^2/df=1.11$ ,  $p=.29$ ; Form 2:  $\chi^2/df=1.12$ ,  $p=.29$ ). Comparison of the means and standard deviations for each item before and after the EM procedure was conducted to showed minimal change to variables. Hence, full information from  $N = 1667$  (Form 1) and  $N = 1501$  (Form 2) was available for further analyses.

In order to examine the effects of self-motivational attitudes towards reading, students were grouped into high-self-efficacy (top third) and low-self-efficacy (bottom third), and high-interest (top third) and low-interest (bottom third). Mean scores per attitude group (Table 3) showed that there were very large mean differences between the groups for self-efficacy (equal to four standard deviations) and for interest (equal to three standard deviations). The scale of difference ensures that 100% of the high groups are above the mean of the lower motivation groups.

Table 3. Mean Motivation Scores by Group and Form

	Total			Form 1			Form 2		
Group	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Self-Efficacy									
High	1040	3.71	0.26	508	3.71	.26	532	3.71	.25
Low	1086	2.27	0.43	572	2.27	.43	514	2.27	.44
Effect Size ( <i>d</i> )		4.03				4.00		4.04	
Interest									
High	1394	3.35	.53	693	3.53	.36	701	3.38	.52
Low	1443	2.70	.62	763	2.70	.62	680	2.69	.63
Effect Size ( <i>d</i> )		1.13			1.62			1.20	

**Model Development.** Exploratory and confirmatory factor analyses of student responses to the two SCoA forms (Form 1  $N=1774$ ; Form 2  $N=1623$ ) was conducted separately, using AMOS v23 (IBM, 2013). No correlated errors were used and Pearson correlations were used. Conventionally, factors are expected to have three or more items to obtain identifiability and consistency (Bandalos & Finney, 2010). However, factors with just two items can be recovered and reported when they are correlated with other factors that are fully identified within the measurement model (Bollen, 1989). Structural equation modelling was then used to establish the relationship of the SCoA measurement models to performance on the asTTle reading tests, on the presumption that beliefs about the purpose of assessment would be adaptive or maladaptive to achievement. Once stable structural models were identified, nested invariance testing between high and low level groups was tested for each form and motivational construct.

While Cronbach alpha estimates of factor reliability are reported, it is noted that these are under-estimates of scale consistency (Sijtsma, 2009). Consequently, this study makes use of the power of confirmatory factor analytic procedures to establish how well the measurement model conforms to the responses in the source data (Hoyle & Duvall, 2004). A range of indexes (Fan & Sivo, 2005; Hu & Bentler, 1999) was used to establish model fit for measurement and structural models because of their different sensitivities to model features such as sample size, model complexity, and model misspecification (Fan & Sivo, 2007; Marsh, Hau, & Wen, 2004). The cut-off values used to indicate acceptable fit were:  $p(\chi^2/df) > .05$ , gamma hat  $> .90$ , SRMR  $\leq .08$  and RMSEA  $\leq .08$ .

**Multi-group Invariance Testing.** Multi-group invariance testing of the structural models was conducted to determine whether the structural model paths linking SCoA to reading performance were invariant for high and low self-efficacy or interest groups (Brown, Harris, O'Quin, & Lane, 2017). Metric and scalar equivalence are required to assume that two groups are drawn from the same population (Wu, Li, & Zumbo, 2007), which permits comparison of their scores. If level of interest and self-efficacy interact with SCoA beliefs, then the regression weights from the various SCoA factors to tested performance on the asTTle reading comprehension should differ by more than chance (i.e., show non-equivalence or non-invariance) and result in meaningful differences in performance. Nested invariance analysis involves using a series of increasingly stringent tests in which specified model parameters are constrained to be equivalent (Cheung & Rensvold, 2002). The sequence of invariance tests followed conventional recommendations:

- The configuration of paths and zero paths had to be identical between groups and is accepted if RMSEA  $\leq .05$ .
- The equivalence of factor to item regression slopes or weights (metric invariance) was accepted if difference in CFI  $\leq .01$ .
- Equivalence of intercepts of the regression slopes at the factor (scalar invariance) was accepted if difference in CFI  $\leq .01$ .

A small change in the comparative fit index ( $\Delta CFI \leq .01$ ) indicates that the introduction of the constraint has not modified the fit of the model to the data by more than chance and so invariance of that constraint can be accepted (Cheung & Rensvold, 2002).

## Results

### Student Conceptions of Assessment Measurement Models

For each form, a measurement model consisting of six factors was found. The six factors were:

- *Student Accountability*: assessment makes me accountable (three identical items in both forms);
- *School Accountability*: assessment makes schools accountable (three identical items in both forms);
- *Enjoyment*: assessment is helpful and enjoyable (two items in Form 1, three in Form 2, with one item common);
- *Informative*: assessment informs me, (two items in Form 1 and four in Form 2, no common items);
- *Unfair*: assessment is frustrating and unfair (three items in both forms, one common item); and
- *Useless*: assessment is useless and worthless (four items in both forms, three common items).

However, these factors were not inter-correlated as first-order factors. Instead hierarchical structure was needed for the positive and negative aspects of student conceptions of assessment. Table 4 shows that Factor 3 (enjoy) was a second-order, superordinate factor which regressed onto two dependent factors (i.e., Factor 1 Student Accountability and 4 Informative). Likewise, Factor 5 (Unfair) was a superordinate factor to the subordinate Factor 6 (Useless). Factor 2 (School Accountability) stood alone. There were correlations among Factors 2, 3, and 5. Hence, each model was multidimensional, hierarchical, and inter-correlated. Each model had good fit characteristics (Form 1:  $n=1531$ ;  $\chi^2=464.03$ ;  $df=113$ ;  $\chi^2/df=4.11$ ,  $p=.04$ ; CFI=.95; gamma hat=.97; RMSEA=.045 (90%CI=.041-.049); SRMR=.041; Form 2:  $n=1435$ ;  $\chi^2=1073.46$ ;  $df=164$ ;  $\chi^2/df=6.55$ ;  $p=.01$ ; CFI=.88; gamma hat=.95; RMSEA=.062 (90%CI=.059-.066); SRMR=.079).

Table 4. Factor Structure for SCoA-II Inventory by Form and level of Self-efficacy and Interest

		Self-efficacy						Interest					
Factor Structure		Loadings		Inter-correlations				Loadings		Inter-correlations			
Superordinate	Sub-factor	High	Low	I.	II.	III.		High	Low	I.	II.	III.	
<i>Form 1</i>													
I. Enjoy				—	0.92	0.43				—	0.92	-0.39	
	IA. Accountable	0.78	0.90					0.81	0.88				
	IB. Informative	0.99	0.97					0.98	0.96				
II. School Quality	—			0.51	—	-0.19				0.95	—	-0.18	
III. Unfair				-0.23	0.04	—				-0.33	-0.21	—	
	IIIA. Useless	0.77	0.77					0.73	0.80				
<i>Form 2</i>													
I. Enjoy				—	0.89	-0.13				—	0.89	-0.13	
	IA. Accountable	0.82	0.95					0.82	0.90				
	IB. Informative	0.93	0.91					0.93	0.95				
II. School Quality	—			0.92	—	0.24				0.94	—	0.24	
III. Unfair				0.43	0.52	—				0.18	0.34	—	
	IIIA. Useless	0.68	0.41					0.68	0.60				

Note. All values from unconstrained model; loadings are standardised beta weights; inter-correlations above diagonal = High group, below diagonal=low group.

Across the two forms, the six factors were given the same labels despite having somewhat different mixes of items. Table 5 provides the items for each factor and form, and scale estimates of reliability, means, and standard deviations. Mean scores for the factors were consistently lower on Form 1 compared to Form 2, but the effect size for these differences ranged from  $|d|=.05$  to  $.32$ , with a mean of  $|d|=.13$ , suggesting that the mean scores for the factors had at best a small difference according to form administered.

Table 5. SCoA II Unconstrained Measurement Model Item Factors, Statements, and Loadings by Form, with Scale Reliability Estimate

Label	Item Statement	Form 1 Loading	Form 2 Loading
<i>Assessment makes me accountable</i>		$\alpha=.69$ ; $M=3.89$ , $SD=0.83$	$\alpha=.66$ ; $M=3.95$ , $SD=0.77$
COAac4	Assessment is assigning a grade or level to my work*	0.67	0.63
COAac5	Assessment is checking off my progress against achievement objectives*	0.68	0.50
COAac6	Assessment is comparing my work against set criteria*	0.55	0.54
<i>Assessment makes schools accountable</i>		$\alpha=.70$ ; $M=3.40$ , $SD=0.99$	$\alpha=.68$ ; $M=3.32$ , $SD=0.98$
COAac8	Assessment keeps schools honest and up-to-scratch*	0.68	0.70
COAac9	Assessment measures the worth or quality of schools*	0.59	0.53
COAac11	Assessment provides information on how well schools are doing*	0.69	0.64
<i>Assessment is helpful &amp; enjoyable</i>		$\alpha=.58$ ; $M=3.23$ , $SD=0.74$	$\alpha=.54$ ; $M=3.29$ , $SD=0.90$
COAimp9	Assessment is an engaging and enjoyable experience for me*	0.77	0.53
COAimp4	Assessment helps me improve my learning	0.55	—
COAimp11	Assessment is integrated with my learning	—	0.52
COAval11	Assessment results predict my future performance	—	0.51
<i>Assessment informs me</i>		$\alpha=.64$ ; $M=3.56$ , $SD=0.67$	$\alpha=.77$ ; $M=3.85$ , $SD=1.12$
COAval3	Assessment identifies how I think	0.65	—
COAval7	Assessment measures my higher order thinking	0.72	—
COAimp12	Assessment makes me do my best	—	0.67
COAimp13	Assessment provides feedback to me about my performance	—	0.64
COAval2	Assessment makes clear and definite what I have learned	—	0.68
COAval9	Assessment results are trustworthy	—	0.60



Assessment is frustrating and unfair		$\alpha=.59$ ; $M=2.80$ , $SD=0.89$	$\alpha=.50$ ; $M=3.11$ , $SD=0.91$
COAir3	Assessment interferes with my learning	0.46	—
COAir5	Assessment is unfair to students	0.76	—
COAir13	Teachers are over-assessing	0.51	0.39
COAir1	Assessment forces me to learn in a way against beliefs about learning	—	0.62
COAir4	Assessment is an imprecise process	—	0.46
Assessment is useless and worthless to me		$\alpha=.72$ ; $M=2.50$ , $SD=1.57$	$\alpha=.71$ ; $M=2.58$ , $SD=1.52$
COAir6	Assessment is value-less	0.61	—
COAir8	I ignore or throw away assessment results*	0.62	0.58
COAir9	I make little use of assessment results*	0.70	0.57
COAir10	I ignore assessment information*	0.52	0.72
COAir2	Assessment has little impact on my learning	—	0.49

Note. Loading values are standardised beta regression weights; \*=items reported previously in Brown and Hirschfeld (2008); items marked '—' were not part of the form.

### SCoA to asTTle Reading Structural Models

Regression paths were introduced from each of the six SCoA factors to asTTle total score. For both forms, the only statistically significant predictor of reading performance was the apparently adaptive conception 'assessment makes me accountable' (Form 1  $\beta=.77^{***}$ , Form 2  $\beta=.99^{***}$ ), while the conception 'assessment is useless' was a statistically significant negative predictor of reading performance (Form 1  $\beta=-.24^{***}$ , Form 2  $\beta=-.15^*$ ). All other paths were statistically not significant due to the inter-correlated nature of the predictors. These structural models had good fit for Form 1 ( $n=1531$ ;  $\chi^2=489.04$ ;  $df=124$ ;  $\chi^2/df=3.94$ ;  $p=.05$ ; CFI=.95; gamma hat=.97; RMSEA=.044 (90%CI=.040-.048); SRMR=.040) and acceptable fit for Form 2 ( $n=1435$ ;  $\chi^2=1172.19$ ;  $df=178$ ;  $\chi^2/df=6.59$ ;  $p=.01$ ; CFI=.87; gamma hat=.94; RMSEA=.062 (90%CI=.059-.066); SRMR=.078). For the purposes of invariance testing between high and low self-efficacy and interest groups, only these statistically significant predictors of reading achievement were used.

### Invariance by Motivation Factors

Table 6 reports the invariance test results for the self-efficacy and interest groups for both forms in the structural models that predict reading achievement. Configural invariance of the structural models was demonstrated across all four comparisons. Metric invariance was found (i.e.,  $\Delta CFI < .01$ ) for all but one comparison (i.e., except self-efficacy in Form 1). This indicates that in three of the four comparisons the regression weights varied by chance, even though the groups differed by significant margins in self-efficacy and interest in reading.

However, scalar invariance was rejected across all comparisons indicating that the high and low groups had different intercepts. This suggests a possible impact explanation; that is, the differing overall means result in a differing intercept value for the regression equation from the latent trait to the manifest variables. In other words, the level of interest and self-efficacy causes a different starting value but not a different strength of relationship to the contributing items.

Table 6. Invariance Test Results for Structural Models by Form and by Interest and Self-efficacy Grouping

Model and Invariance Test	Self-efficacy			Interest		
	RMSEA	CFI	$\Delta CFI$	RMSEA	CFI	$\Delta CFI$
Form 1 Structural Model with Reading						
Test 1 Configural	.050	.860		.038	.924	
Test 2 Metric (weak)	.054	.829	.031	.038	.919	<b>.005</b>
Test 3 Scalar (strong)	.058	.791	.038	.043	.886	.033
Form 2 Structural Model with Reading						
Test 1 Configural	.045	.860		.045	.859	
Test 2 Metric (weak)	.045	.855	<b>.005</b>	.045	.853	<b>.006</b>
Test 3 Scalar (strong)	.053	.791	.064	.050	.805	.048

Note. RMSEA = root mean square error of approximation; CFI = comparative fit index; values in bold represent equivalent fit after constraint imposed.

The structural model regression weights were examined for each group to identify how the groups differed (Table 7). There were large observable differences in regression weights between the High and Low motivation groups in the unconstrained condition. However, after constraining the proven metric equivalence, the regression weights differed by a small amount (i.e.,  $<.05$ ) or were within chance. Once constrained the amount of variance in reading achievement explained by these conceptions of assessment was almost identical and small. This indicates that constraining the models for metric equivalence resulted in effects that were fundamentally equivalent for both low and high groups.

The exception to this result was the difference between high and low interest in Form 1 only, which failed to reach metric equivalence. The difference of effect for the factor student accountability was negative for the high group and not statistically significant for the low group. The assessment is useless factor was negative for both high and low groups, though more so for the high interest group. However, inspection of 95%CI from maximum likelihood bootstrap values based on 1000 samples suggested there was a small overlap in the path values for assessment makes me accountable, and no overlap for assessment is useless.

Table 7. Regression Weight of Student Year upon SCoA and Reading Achievement by Form and Level of Interest or Self-efficacy

	Student Conceptions of Assessment Factors				Variance in reading achievement ( $R^2$ )	
	<u>Makes me accountable</u>		<u>Useless &amp; worthless</u>			
Motivation by Form	High	Low	High	Low	High	Low
<i>Self-efficacy</i>						
Form 1 (unconstrained)	-0.17**	0.08	-0.33***	-0.04	0.11	0.01
Metric Invariance	-0.03	-0.03	-0.21***	-0.16***	0.04	0.03
Form 2 (unconstrained)	-0.15**	0.04	-0.27***	-0.18**	0.09	0.03
Metric Invariance	-0.05	-0.05	-0.22***	-0.19***	0.05	0.04
<i>Interest</i>						
Form 1 (unconstrained)	-0.21***	0.01	-0.30***	-0.09*	0.10	0.01
Metric Invariance	—	—	—	—	—	—
Bootstrap 95%CI	-0.35 - 0.08**	-0.09 - 0.11	-0.40 - 0.20**	-0.18 - 0.00	0.05 - 0.19	0.001 - 0.04
Form 2 (unconstrained)	-0.20***	0.09*	-0.28***	-0.14**	0.11	0.02
Metric Invariance	-0.04	-0.04	-0.18***	-0.19***	0.03	0.04

Note: Values are standardised beta weights; \* $p<.05$ , \*\* $p<.01$ , \*\*\* $p<.001$

Because years of schooling clearly mattered to achievement in reading, this variable was introduced into the two structural equation models to examine whether student year also mattered to the structural model and its invariance testing. Table 8 shows the regression weight of student year to the three key student conceptions of assessment and reading score by level of interest and self-efficacy according to SCoA Form. The data show that while year matters to reading, it has a statistically significant effect on conceptions of assessment in just 6 of the 24 possible relationships. The strength of the statistically significant paths was small ( $|\beta|=.11-.14$ ) and there did not appear to be a pattern according to high or low group. There were equal numbers of statistically significant paths for high and low groups. However, five of the six paths are found in Form 1, with three on assessment makes me accountable (low Interest and Self-efficacy and High Self-efficacy) and two on assessment is enjoyable (Low Interest and High Self-Efficacy). Interestingly, all five paths were negative, indicating that as student year increased, the conception of assessment as enjoyable or making me accountable decreased. The most important conclusion from this analysis is that student year is largely independent of how conceptions of assessment are formed and how it interacts with levels of interest and self-efficacy.

Table 8: SCoA Factor Loadings on Reading Achievement by Form and Group for Statistically Significant Predictors of Achievement

	Student Conceptions of Assessment			Reading Score
	Unfair	Joy	Accountability	
<i>Form 1</i>				
High Interest	0.08 <sup>ns</sup>	-0.08 <sup>ns</sup>	-0.10 <sup>ns</sup>	0.41***
Low Interest	0.02 <sup>ns</sup>	-0.09*	-0.11*	0.62***
High Self-Efficacy	-.01 <sup>ns</sup>	-0.11*	-0.12*	0.43***
Low Self-Efficacy	-.05 <sup>ns</sup>	-0.07 <sup>ns</sup>	-0.11*	0.63***
<i>Form 2</i>				
High Interest	0.04 <sup>ns</sup>	0.01 <sup>ns</sup>	-0.03 <sup>ns</sup>	0.68***
Low Interest	0.06 <sup>ns</sup>	-0.01 <sup>ns</sup>	0.00 <sup>ns</sup>	0.53***
High Self-Efficacy	0.14*	-0.06 <sup>ns</sup>	-0.02 <sup>ns</sup>	0.53***
Low Self-Efficacy	-0.02 <sup>ns</sup>	0.00 <sup>ns</sup>	-0.07 <sup>ns</sup>	0.66***

Note. Values are standardised beta regression weights; ns=not statistically significant; \*= $p < .05$ ; \*\*= $p < .01$ ; \*\*\*= $p < .001$

## Discussion

This study demonstrates that students' achievement in reading is impacted in a small way by their conceptions of assessment. Specifically, two conceptions (i.e., assessment makes me accountable and assessment is useless) had statistically significant relations to achievement. While unconstrained values were different in direction and scale for assessment makes me accountable, under metric equivalence this difference disappeared and the path values became statistically not significant. In contrast, the effect of assessment is useless was consistently negative, to almost the same degree in both groups, under conditions of metric equivalence. Hence, this study, in three of the four conditions, rejected the hypothesis that these conceptions of assessment have inverse effects for high and low self-efficacy and interest groups. The exception (i.e., assessment is useless) was more strongly negative for the high group than the low group, but the direction of effect was similar (i.e., both negative). The regression slopes, while weak, were equivalent, meaning the same adaptive or maladaptive effects on reading achievement are seen for conceptions of assessment regardless of level of interest or self-efficacy.

The lack of scalar invariance is consistent, as per design of the study, with the quite different mean values seen in the high and low groups. This suggests that the real-world difference between these two groups of students had impact on their responses to the items. When the regression weights of traits to items are equivalent, it would seem that the intercept values on the latent trait have to be different to account for the different means.

Thus, this study shows that having high or low level of interest and self-efficacy makes little difference in how conceptions of assessment influence achievement on a standardised reading test. Unlike a previous study (Brown & Hirschfeld, 2008), the influence of 'assessment makes me accountable' was statistically not significant when taking into account student interest or self-efficacy in reading. In contrast, assessment is useless had a consistently negative impact on achievement, which is consistent with the same study (Brown & Hirschfeld, 2008). It is plausible that endorsing the uselessness of assessment reflects the antithesis of SRL. SRL suggests that self-reflection about achieved outcomes is an integral part of using adaptive learning. Certainly, the negative pathway to achievement from this conception indicates that it is a maladaptive approach to assessment.

The size of impact of these conceptions of assessment upon test achievement was small ( $R^2 \leq .05$ ). This means that an increase of one *SD*, especially in the assessment is useless factor would theoretically produce a change of up to five points on the asTTle version 4 reading comprehension test. It may be that this belief has a statistically significant impact on scores among New Zealand high school students because the classroom teaching practices they experience do not necessarily turn assessment events or results into useful feedback about how to improve (Brown, Irving, Peterson, & Hirschfeld, 2009). Nevertheless, a small increase, especially if it were relatively easy to persuade students to move away from the conception that assessment is useless, would be a worthwhile objective.

Interestingly, only two of the six conceptions of assessment had a statistically significant direct effect on achievement. Later versions of the Student Conceptions of Assessment (Brown,

Peterson, & Irving, 2009), built partly on this study, have reported that both direct and indirect paths exist from student conceptions of assessment to performance in mathematics. The discrepancy in results may partly arise from the difference of subject matter (i.e., mathematics is not the same as reading).

Although this study's structural equation model suggests causal pathways (i.e., conceptions of assessment cause achievement results), this study, making use of cross-sectional data, cannot establish such claims. Nonetheless, since the models were reasonably similar across forms and because adaptive and maladaptive paths are consistent with SRL theory, the results do suggest testable hypotheses. These hypotheses can only be validated in longitudinal intervention studies that attempt to modify student belief systems about assessment (see suggested studies described in Brown, McInerney, & Liem, 2009).

Nonetheless, this study suggests that helping students to think in a more self-regulatory fashion about assessment will have similar amounts of positive effect on achievement for students with either high or low levels of self-efficacy and interest in the subject. The lack of difference between high and low self-efficacy for how conceptions of assessment relate to achievement does not mean that having greater self-efficacy is unimportant. That the path values to achievement were largely equivalent suggests that increasing self-efficacy will benefit all students. It may be that research into student interest and self-efficacy would benefit from incorporating attention to how students regulate their responses to assessment events, processes, and results. Clearly, SRL requires self-efficacy in not just learning but also in adaptively controlling beliefs and strategies about assessment in preparation for, during assessment administration, and in reflection upon actual assessment results (McMillan, 2016).

Believing that assessment is a valid evaluation and description of performance and acting on that belief to improve learning is likely to benefit SRL and learning outcomes. Being self-efficacious for assessment processes themselves is also likely to contribute to greater outcomes. Hence, if all students can be helped to embrace adaptive values concerning assessment (i.e., see assessment as useful in helping them evaluate and improve their progress towards valued goals, and stop making excuses for poor achievement) there are potentially useful pay offs in terms of their academic progress. For these students changing both their motivational attitudes and their attitudes towards assessment appears to be a promising strategy.

## Acknowledgements

The authors wish to thank Professor John Hattie for access to the asTTle database from which these data were drawn. The opinions expressed here are those of the authors and not of the Ministry of Education which funded the asTTle development. An earlier version of this paper was presented in Walton, K. F. (2009). Secondary students' conceptions of assessment mediated by self-motivational attitudes: Effects on academic performance. Unpublished master's thesis, University of Auckland, Auckland, NZ, supervised jointly by Dr Gavin T. L. Brown and Dr Susan Farruggia.

## References

- Alexander, P. A. (1995). Superimposing a situation-specific and domain-specific perspective on an account of.. *Educational Psychologist*, 30(4), 189-193. doi:10.1207/s15326985ep3004\_3
- Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G. R. Hancock & R. O. Mueller (Eds.), *The Reviewer's Guide to Quantitative Methods in the Social Sciences* (pp. 93-114). New York: Routledge.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, 78(1), 246-263. doi:10.1111/j.1467-8624.2007.00995.x
- Boekaerts, M. (1995). Self-regulated learning: Bridging the gap between metacognitive and metamotivation theories. *Educational Psychologist*, 30(4), 195-200.
- Boekaerts, M., & Cascallar, E. (2006). How far have we moved toward the integration of theory and practice in self-regulation? *Educational Psychology Review*, 18, 199-210.
- 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.
- Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: John Wiley & Sons, Inc.
- Bong, M. (2013). Self-efficacy. In J. Hattie & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 64-66). New York: Routledge.
- Brown, G. T. L. (2004). Measuring attitude with positively packed self-report ratings: Comparison of agreement and frequency scales. *Psychological Reports*, 94(3), 1015-1024. doi:10.2466/pr0.94.3.1015-1024
- Brown, G. T. L. (2011). Self-regulation of assessment beliefs and attitudes: A review of the Students' Conceptions of Assessment inventory. *Educational Psychology*, 31(6), 731-748. doi:10.1080/01443410.2011.599836
- Brown, G. T. L. (2013). Student conceptions of assessment across cultural and contextual differences: University student perspectives of assessment from Brazil, China, Hong Kong, and New Zealand. In G.A.D. Liem & A. B. I. Bernardo (Eds.), *Advancing Cross-cultural Perspectives on Educational Psychology: A Festschrift for Dennis McInerney* (pp. 143-167). Charlotte, NC: Information Age Publishing.
- Brown, G. T. L., & Hirschfeld, G. H. F. (2005, December). Secondary school students' conceptions of assessment. *Conceptions of Assessment and Feedback Project Report #4. Auckland: University of Auckland*. doi:10.13140/RG.2.2.11541.93921
- Brown, G. T. L., & Hirschfeld, G. H. F. (2007). Students' conceptions of assessment and mathematics: Self-regulation raises achievement. *Australian Journal of Educational and Developmental Psychology*, 7, 63-74.
- Brown, G. T. L., & Hirschfeld, G. H. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy and Practice*, 15(1), 3-17. doi:10.1080/09695940701876003
- Brown, G. T. L., Harris, L. R., O'Quin, C. R., & Lane, K. (2017). Using multi-group confirmatory factor analysis to evaluate cross-cultural research: Identifying and understanding non-invariance. *International Journal of Research and Method in Education*, 40(1), 66-90. doi:10.1080/1743727X.2015.1070823
- 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. doi:10.1016/j.learninstruc.2008.02.003
- Brown, G. T. L., Peterson, E. R., & Irving, S. E. (2009). Beliefs that make a difference: Adaptive and maladaptive self-regulation in students' 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. 159-186). Charlotte, NC: Information Age Publishing.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233-255.
- Comparative Education Research Unit. (2016). New Zealand Headline Results: PISA 2015. Wellington, NZ: Ministry of Education. Retrieved from



- [https://www.educationcounts.govt.nz/\\_data/assets/pdf\\_file/0003/180597/PISA-2015-Headlines-v2.pdf](https://www.educationcounts.govt.nz/_data/assets/pdf_file/0003/180597/PISA-2015-Headlines-v2.pdf)
- 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.
- 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.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah N.J.: Lawrence Erlbaum Associates.
- Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Models*, 12, 343-367.
- Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, 42(3), 509-529.
- Harris, L. R., & Brown, G. T. L. (2013). Opportunities and obstacles to consider when using peer- and self-assessment to improve student learning: Case studies into teachers' implementation. *Teaching and Teacher Education*, 36, 101-111. doi:10.1016/j.tate.2013.07.008
- Hattie, J. A. (2004). *Models of self-concept that are neither top-down or bottom-up: The rope model of self-concept*. Paper presented at the 3rd International Biennial Self Research, Berlin.
- Hattie, J. A., Brown, G. T. L., & Keegan, P. J. (2003). A national teacher-managed, curriculum-based assessment system: Assessment tools for teaching and learning (asTTle). *International Journal of Learning*, 10, 771-778.
- Hattie, J. A., Brown, G. T. L., Keegan, P. J., MacKay, A. J., Irving, S. E., Cutforth, S., et al. (2004). *Assessment Tools for Teaching and Learning (asTTle) Version 4, 2005: Manual*. Wellington: University of Auckland/ Ministry of Education/ Learning Media.
- Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research*, 70(2), 151-179.
- 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. doi:10.1027/1015-5759.25.1.30
- Hoyle, R. H., & Duvall, J. L. (2004). Determining the number of factors in exploratory and confirmatory factor analysis. In D. Kaplan (Ed.), *The SAGE Handbook of Quantitative Methodology for Social Sciences* (pp. 301-315). Thousand Oaks, CA: Sage.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1-55.
- IBM. (2013). AMOS [computer program] (Version 23, Build 1812). Wexford, PA: Amos Development Corporation.
- 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.
- Marsh, H. W., Hau, K., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers of overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, 11(3), 320-341.
- Marsh, H., Hau, K., Artelt, C., Baumert, J., & Peschar, J. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing*, 6(4), 311-360.
- McMillan, J. H. (2016). Section discussion: Student perceptions of assessment. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Human and Social Conditions in Assessment* (pp. 221-243). New York: Routledge.
- Meyer, L. H., McClure, J., Walkey, F., Weir, K. F., & McKenzie, L. (2009). Secondary student motivation orientations and standards-based achievement outcomes. *British Journal of Educational Psychology*, 79(2), 273-293. doi:10.1348/000709908X354591
- Ministry of Education. (2007). *The New Zealand curriculum for English-medium teaching and learning in years 1-13*. Wellington: Learning Media Ltd.

- 'Otunuku, M., & Brown, G. T. L. (2007). Tongan students' attitudes towards their subjects in New Zealand relative to their academic achievement. *Asia Pacific Education Review*, 8(1), 117-128. doi:10.1007/BF03025838
- 'Otunuku, M., Brown, G. T. L., & Airini. (2013). Tongan secondary students' conceptions of schooling in New Zealand relative to their academic achievement. *Asia Pacific Education Review*, 14(3), 345-357. doi:10.1007/s12564-013-9264-y
- Pajares, M. F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543-578.
- Schunk, D. H. (1983). Progress self-monitoring: Effects on children's self-efficacy and achievement. *Journal of Experimental Education*, 51, 89-93.
- Schunk, D. H., & Ertmer, P. A. (2000). Self-regulation and academic learning: self-efficacy enhancing interventions. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 631-649). San Diego: Academic Press.
- Schunk, D. H., & Zimmerman, B. J. (2006). Competence and control beliefs: Distinguishing the means and ends. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of Educational Psychology* (2nd ed., pp. 349-367). Mahwah, NJ: LEA.
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, 74(1), 107-120. 10.1007/S11336-008-9101-0
- Tanaka, J. S. (1987). "How Big Is Big Enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child Development*, 58(1), 134-146.
- Walton, K. F. (2009). *Secondary students' conceptions of assessment mediated by self-motivational attitudes: Effects on academic performance*. (Unpublished M.Ed. thesis), University of Auckland, Auckland, NZ.
- Weiner, B. (1986). *An attribution theory of motivation and emotion*. New York: Springer-Verlag.
- Weiner, B. (2000). Intrapersonal and interpersonal theories of motivation from an attributional perspective. *Educational Psychology Review*, 12, 1-14.
- Wise, S. L., & Cotten, M. R. (2009). Test-taking effort and score validity: The influence of student 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. 187-205). Charlotte, NC: InformationAge Publishing.
- Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation*, 12(3), 1-26.
- Zimmerman, B. J. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed., pp. 1-37). Mahwah, NJ: LEA.
- Zimmerman, B. J., & Schunk, D. H. (2004). Self-regulating intellectual processes and outcomes: A social cognitive perspective. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion and cognition*. Mahwah, NJ: Lawrence Erlbaum Associates.