

Formative and summative assessment in Hong Kong primary schools: students' attitudes matter

Wu Yuan Guo & Zi Yan

To cite this article: Wu Yuan Guo & Zi Yan (2019) Formative and summative assessment in Hong Kong primary schools: students' attitudes matter, *Assessment in Education: Principles, Policy & Practice*, 26:6, 675-699, DOI: [10.1080/0969594X.2019.1571993](https://doi.org/10.1080/0969594X.2019.1571993)

To link to this article: <https://doi.org/10.1080/0969594X.2019.1571993>



Published online: 15 Feb 2019.



Submit your article to this journal [↗](#)



Article views: 515



View related articles [↗](#)



View Crossmark data [↗](#)



Formative and summative assessment in Hong Kong primary schools: students' attitudes matter

Wu Yuan Guo  and Zi Yan 

Department of Curriculum and Instruction, The Education University of Hong Kong, Hong Kong, China

ABSTRACT

This study investigates the relationship between students' attitudes towards formative assessment and summative assessment and aims to enrich the understanding of formative and summative assessment from the students' perspective. A total of 3,019 Hong Kong primary school students responded to a newly developed instrument. Overall, students reported positive instrumental yet negative affective attitudes towards formative and summative assessment. Girls had less negative affective attitudes towards formative assessment, but more positive instrumental attitudes towards formative and summative assessment than boys. Grade 4 students consistently scored higher than Grade 5 and Grade 6 students in assessment attitudes. The gender and grade differences were statistically significant, but had small effect size. Moreover, it was found that students' affective and instrumental attitudes to formative assessment positively predicted students' affective and instrumental attitudes to summative assessment.

ARTICLE HISTORY

Received 21 February 2018
Accepted 5 January 2019

KEYWORDS

Formative assessment;
summative assessment;
instrumental attitude;
affective attitude; primary
school

Introduction

Global education reforms have emphasised that the ultimate goal of education is not to evaluate students, but to help them become lifelong learners (Bransford, Brown, & Cocking, 2001). Traditionally, summative assessment (SA) – i.e. assessment *of* learning – has played a dominant role in educational assessment to evaluate students' academic performances and outcomes. However, formative assessment (FA) – i.e. assessment *for* learning – has gradually become more recognised as helping students to improve learning by providing ongoing feedback during the learning process (Kwon, Lee, & Shin, 2017). Apparently, the primary factor that differentiates SA and FA is purpose. SA is used to make final judgements about students' academic performance and is often instantiated as high-stakes examinations for educational selection and accountability purposes (Glover, Reddy, Kettler, Kurz, & Lekwa, 2016), which is especially dominant in an exam-oriented culture such as Hong Kong (Choi, 1999; Jackson, 2013; Law, 2007). FA is utilised mainly to inculcate students with self-regulated learning skills to allow students to become the real masters of their own study (Clark, 2012; Yan & Cheng, 2015). In line with the global educational trend to cultivate lifelong learners, the Hong Kong government has been working for decades in an effort to promote FA in schools (Curriculum Development

Council, 2001). The Education Bureau documents explicitly state that ‘there should be a change in the assessment practices and schools should put more emphasis on “Assessment *for* Learning” as an integral part of the learning, teaching and assessment cycle.’ (Curriculum Development Council, 2001).

However, such a change in assessment culture is by no means an easy task. FA is often interpreted very seriously, yet incorrectly, by schools, teachers, students, and parents. Although FA is advocated by Hong Kong government policy and intended to balance excessive testing, it has not been well executed in Hong Kong schools given the dominance of SA (Berry, 2011; Lee, 2007; Yan & Cheng, 2015). A profound reason that might largely account for this phenomenon in Hong Kong is children, teachers and parents’ limited knowledge of FA, in addition to a stereotypical view of the nature of assessment. Therefore, in order to help these various stakeholders better accept FA, their perceptions of assessment – especially FA – need to be thoroughly understood. Among all the stakeholders involved in FA, students – the most direct participants – deserve particular attention.

FA and SA have often been treated as two isolated assessments. However, some researchers (Biggs, 1998; Roos & Hamilton, 2005; Taras, 2005) refute the view that the two assessments are branches of two different trees. Overlapping functions between FA and SA have been discovered, and interactive uses of the two assessments have been practiced (Carless, 2011; Hui, Brown, & Chan, 2017; Roos & Hamilton, 2005). Both FA and SA can serve a meaningful purpose for education (Brady & Kennedy, 2005). Therefore, the current study argues that if FA and SA can be systematically applied in students’ learning, the reciprocal effects of both assessments could be optimised. How to combine these two powerful forms of assessment to enhance learning processes is a meaningful issue worthy of investigation. Therefore, this research delves into the possible relationships between students’ attitudes towards FA and their attitudes towards SA.

Based on current assessment practices in Hong Kong, this study has a dual-purpose. First, it aims to investigate Hong Kong primary school students’ attitudes towards FA and SA and to examine whether there are differences in attitudes for two key demographic values (i.e. gender and grade) as mixed findings have been reported regarding students’ attitudes to assessment for each gender and grade level (Alkharusi, 2013; Awofala & Babajide, 2013; Berry, 2011; Brown & Harris, 2012; Davis & Brember, 1998; Dhindsa, Omar, & Waldrip, 2007; Lee, 2007; McMillan, 2016; Meece, Herman, & McCombs, 2003; Yan, 2018). Second, although FA and SA are distinguishable from each other in many aspects, they should not be simply treated as separate; thus, the association between attitudes towards FA and SA is investigated.

Understanding formative and summative assessment

In spite of the fact that the fundamental focus of FA is assessment *for* learning, many researchers have defined FA using different conceptualisations. Black and William (1998) stated that FA is a kind of assessment that encompasses all activities undertaken by teachers and students, which provides feedback to calibrate the teaching or learning process. This perspective stresses that FA serves learning and could be used on a daily basis. Heritage (2007) further purported four components of FA: 1) Identify the ‘right gap’ in learning. 2) Provide instant feedback for students’ learning. 3) Involve students in the assessment activity. 4) Monitor and regulate students’ learning progress.

In sum, FA is an ongoing process that helps students to regulate their own learning. In the current study, FA is operationally defined as an assessment that is utilised to provide students with feedback, facilitate students' learning, and check their academic progress (Sadler, 1998). Examples of FA might include tests/worksheets implemented by teachers in class or informal quizzes initiated by students or their peers, with instant feedback to check learning progress.

Previous research studies have shown that there are multiple benefits of FA for students. According to Cauley and McMillan's (2010) study, FA plays a powerful role in enhancing students' academic motivation and performance. Passion for life-long learning is instilled in students by exercising FA (Young & Jackman, 2014). It was reported that speed of learning could be significantly raised among learners who use formative techniques to assess their performance (Volante & Beckett, 2011). Moreover, FA helps students take control of their study and eventually become self-regulated learners (Nicol & Macfarlane-Dick, 2006; Panadero & Jonsson, 2013), which strongly echoes the revolutionary change in assessment strategy reform. On the basis of all empirical evidence, students can benefit substantively from FA on a regular basis.

In contrast with FA which focuses on the learning process, SA highlights the final performance in which the evaluative role of SA clearly stands out. On one hand, Black and William (1998) concluded that SA refers to official, systematic, and high-stakes examinations for educational evaluation, selection, and decision purposes. SA is widely utilised to screen and select students for the next level of education or employment. On the other hand, SA is naturally used to evaluate and summarise students' learning performance. Taras (2005) perceived SA as a judgmental point at which all performance evidence is conclusively gathered. Operationally, in this study, SA refers to the exams that are developed by assessment authorities (e.g. teachers, exam experts) to test students' learning outcomes. These exams usually take place at the end of a particular learning period, such as mid-semester or end-of-semester examinations.

Alongside the dominant role of SA, overuse of SA gradually became accepted as 'normal' in schools in Hong Kong (Choi, 1999; Jackson, 2013; Law, 2007) as well as in many other countries (Marinho, Leite, & Fernandes, 2017; Rust, 2000). Although SA serves as a more impartial and reliable approach to assessing students' academic competences, it does have some negative effects. Due to its often high-stakes nature, students are faced with a heavy workload and psychological stress in preparation for SA (Choi, 1999; Law, 2007). Knight (2002) found surface learning approaches (students trying to learn by memorising and reproducing factual knowledge) are often the consequences of SA, which harm students' learning quality. All these undesired aspects go against the ultimate educational goal of cultivating lifelong learners. Therefore, the Hong Kong government decreed that 'there should be a change in the assessment school practice' (Curriculum Development Council, 2001).

Students' attitudes towards summative and formative assessment

Despite the necessity and value of SA in educational evaluation, there have been mixed findings regarding students' attitudes towards SA. Some students agree that SA is objective, reliable, and relatively stable in assessing their academic performance (Bacon & Bean, 2006). However, SA is also perceived negatively by students because

it is usually presented in the form of high-stakes examinations which could cause test anxiety (Deb, Strodl, & Sun, 2014; Eady, 1999; Gregor, 2005; Reay & Wiliam, 1999). Carless and Lam (2014) also reported in a qualitative study that Hong Kong primary school students had negative feelings about testing.

In order to maximise the optimal effect of FA on student learning, it is important to understand students' attitudes towards FA. Conlon (2006) pointed out that students' attitudes towards FA seem to be essential for investigating the potential contribution of FA. There are some empirical studies in regard to Hong Kong students' conceptions of assessment (Brown & Wang, 2016; Hue, Leung, & Kennedy, 2015; Lee, 2007) which found that students dislike assessment, yet see it as a tool for their study. However, very few studies focus on assessment of primary school students (Carless & Lam, 2014). Carless and Lam (2014) took a qualitative approach to understanding assessment status at the primary school stage and called for more research studies to explore primary school students' attitudes towards assessment. Therefore, primary school students' attitudes towards FA merit in-depth study.

In general, students tend to exhibit mixed attitudes towards assessment (Brown & Harris, 2016; McMillan, 2016). On one hand, student's perceptions may possibly vary towards formative or classroom assessment versus summative or high-stake assessment. On the other hand, both positive and negative feelings could mutually exist towards assessment. As pointed out by McMillan (2016) that 'it is possible to both value something and dislike it, while it is also possible to believe something is important and enjoyable.' Research looking at both views is highly recommended in the educational assessment field.

Mixed findings are also found in gender and grade differences in terms of student's attitude to assessment (Alkharusi, 2013; Awofala & Babajide, 2013; Berry, 2011; Brown & Harris, 2012; Davis & Brember, 1998; Dhindsa et al., 2007; Lee, 2007; McMillan, 2016; Meece et al., 2003; Yan, 2018). Some researchers (Awofala & Babajide, 2013; Conlon, 2006) argued that gender and age (equivalent to grade) were not the key factors that affect college student's attitude to their assessment practices. Dhindsa et al. (2007) had similar results by reporting low effect size in gender and grade differences which indicated these two variables were of little influence to secondary student's assessment attitude. Nevertheless, McMillan (2016) concluded that a student's perception of assessment would vary according to age. His claim was supported by multiple empirical studies showing that girls (Alkharusi, 2013; Yan, 2018; Meece et al., 2003) and lower grade students (Berry, 2011; Brown & Harris, 2012; Davis & Brember, 1998; Lee, 2007) perceived assessment more positively than boys and higher grade students. Due to inconsistent findings, further empirical research on gender and grade can contribute to testify the evidence of such differences.

The relationship between students' attitudes towards formative and summative assessment

In the context of FA being the advocated practice, will students' attitudes towards SA change accordingly? Will students' attitudes to SA be influenced by their attitudes to FA, or vice versa? As Biggs (1998) pointed out, 'the powerful interaction between FA and SA could usefully be incorporated in an overall synthesis, so that both SA and FA are

conceptualised within the same framework'. Thus, if we want to place FA and SA in harmony and help students overcome the negative effects of SA, it's necessary to understand the relationship between students' attitudes towards the two types of assessment.

The connection and interaction between FA and SA have been introduced and discussed by some researchers (Biggs, 1998; Hui et al., 2017; Roos & Hamilton, 2005). Although it is very important to understand and resolve the tension between FA and SA (Brown & Kennedy et al., 2009), very few studies exist which focus on that topic, especially from the students' perspective.

This study

The determinant role of attitude in human behaviour is highlighted in psychological theories, such as the Theory of Planned Behaviour (TPB) proposed by Ajzen (1985, 1991). Research (Fishbein & Ajzen, 1975) shows that behaviour is best predicted by an individual's attitude towards that behaviour. In a review of studies applying TPB (Armitage & Conner, 2001), attitude was recognised as the most influential predictor of behaviours among all the constructs in TPB. In the Hong Kong context, Yan and Cheng (2015) investigated teachers' attitudes regarding FA from the TPB perspective. They found that teachers' instrumental attitudes to FA were an effective predictor for teachers' intentions to practise FA in the classroom. By the same token, in order to know if students would actually engage in assessments and gain learning benefits from such an engagement, it is necessary to understand students' attitudes to those assessments.

According to Ajzen (1985, 1991), there are two types of attitudes: affective attitude and instrumental attitude. Affective attitude refers to enjoyable or unenjoyable emotions and feelings attached to the occurrence of a behaviour. Instrumental attitude incorporates an individual's consequential appraisal of a behaviour as being useful or harmful. When individuals possess high and positive affective and instrumental attitudes towards FA, they are more likely to conduct FA and improve their learning processes as well as outcomes. Similarly, when students perceive SA in a negative way, they tend to avoid SA as much as possible. In alignment with the purpose of the current study, the authors framed this study from the TPB – particularly the attitude – perspective.

Given the literature review that mixed findings about the conception of assessment have been reported with regard to demographic variables such as gender and grade, this study aims to contribute further evidence to clarify the differences of gender and grade in student's assessment attitude.

Overall, this study answers the following questions:

- (1) What are the affective and instrumental attitudes of Hong Kong primary school students towards FA and SA?
- (2) Are there gender differences in Hong Kong primary school students' affective and instrumental attitudes towards FA and SA?
- (3) Are there grade differences in Hong Kong primary school students' affective and instrumental attitudes towards FA and SA?
- (4) Do students' affective and instrumental attitudes towards FA predict their affective and instrumental attitudes towards SA?

Method

Participants

The participants consisted of 3,019 students from 10 Hong Kong primary schools. Considering the language literacy and cognition capability necessary for assessment, only Year 4–6 students aged from 9 to 12 years old were recruited. A cluster sampling method, using school as the unit instead of randomly sampling individual students, was applied because the former approach is more feasible and practical for large sample size studies such as this one (Cohen, Manion, & Morrison, 2011). Participants anonymously responded to the self-developed 30-item instrument which was distributed and collected by teachers in schools. Of the 3,019 participants, there were 1,558 (51.6%) males, 1,427 (47.3%) females, and 34 (1.1%) participants with missing gender information. The sample consisted of 889 (29.4%), 992 (32.9%), and 1,126 (37.3%) Grade 4, 5, and 6 students, respectively, and 12 (0.4%) participants did not provide grade information.

Instrument

There are some established instruments designed to measure different parties' conceptions of assessment. Pat-EI, Tillema, Segers, and Vedder's (2015) survey focused on the congruency of student and teacher perceptions of FA. Several versions of the Students' Conceptions of Assessment Inventory have been developed and validated (Brown, 2011; Brown & Hirschfeld, 2008). These inventories are effective tools to measure students' conceptions of assessment in general. Despite of the existence of available instruments, there is still a need to develop a new instrument for the present study. A core aim of the present study is to differentiate students' attitudes towards assessment into affective and instrumental attitudes according to TPB. However, none of the available instruments make such a distinction. By gauging affective (Is it enjoyable or not?) and instrumental (Is it useful or not?) aspects, a rich picture of students' attitudes towards assessment will be provided.

Based on the reasons above, the authors developed an instrument in Chinese to measure students' attitudes towards FA and SA (see [Appendix 1](#)). The original item pool was drawn from available instruments that assess relevant constructs, such as teachers' or students' perceptions of assessment (Brown, 2011; Brown & Irving et al., 2009; Brown & Wang, 2016; Yan, 2016; Yan & Cheng, 2015), as well as two focus group interviews with primary school teachers ($N = 5$) and students ($N = 12$). The original item pool had 42 items with 21 for affective attitude and 21 for instrumental attitude. The items in the FA and SA domains corresponded to each other. All items were subject to a comprehensive face and content validity check by an expert panel that consisted of three professors in the educational assessment research area and two experienced primary school teachers. Ten items were removed and the refined instrument consisted of four unidimensional scales with 32 items, including a seven-item scale for affective attitude towards FA (AAFA; e.g. I like FA), a nine-item scale for instrumental attitude towards FA (IAFA; e.g. FA can increase my learning motivation), a seven-item scale for affective attitude towards SA (AASA; e.g. I like SA), and a nine-item scale for instrumental attitude towards SA (IASA; e.g. SA can increase my learning motivation). The operational definitions of FA and SA were explained at the beginning

of the instrument and a set of six ordered response options (1 = totally disagree to 6 = totally agree) was provided for each item. Prior to being distributed to the 3,019 participants, the 32-item instrument (see [Appendix 1](#)) was responded to by 283 primary school students as a pilot study in order to examine scale psychometric properties.

Data analysis

The Rasch model analysis (Rasch, 1960), descriptive and inferential statistics, and hierarchical regression analysis were applied in the present study. Firstly, a Rasch rating scale analysis by Winsteps 3.7 (Linacre, 2011) was used to examine the psychometric properties of the scale as well as to provide person measures which are to be used in subsequent analyses (Bond & Fox, 2015). The Rasch model has been used in a similar manner in many studies (e.g. Deneen, Brown, Bond, & Shroff, 2013; Yan, 2014, 2016, 2017). Rasch analysis can transform raw data, collected using a Likert-type scale, into interval data which is a pre-requisite for conventional statistical analysis, such as regression analysis (Bond & Fox, 2015; Wright, 1997). For the purpose of checking the psychometric quality of an instrument, Rasch analysis can generate multiple indices, such as Rasch reliability, item fit statistics, and variance explained by Rasch measures.

After the development and validation of the instrument using a sample of 283 students, the Rasch-calibrated person measures of the main sample ($N = 3,019$) were then subject to further analyses with regard to gender and grade differences in attitude towards FA and SA. Subsequently, hierarchical regression analysis was used to investigate whether students' attitudes to FA could predict their attitudes to SA. In psychological and educational data analysis, hierarchical regression analysis has been widely used to predict the relationship between predictor and dependent variables (Montgomery, Peck, & Vining, 2015).

Results

Psychometric properties of the instrument

A sample ($N = 283$) of Hong Kong Grades 4–6 students was used in the pilot study to examine psychometric properties of the scales. According to Linacre (2011), the MNSQ ranges out of 0.5–1.5 in both Infit and Outfit indicated that the items did not fit the unidimensional scale well. One reversely worded item (FA/SA is a burden for my learning) in the instrumental attitude domain had a MNSQ value above 1.5 (For FA: Infit MNSQ = 1.66/Outfit MNSQ = 1.83; For SA: Infit MNSQ = 1.60/Outfit MNSQ = 2.33); hence, this item was removed from both the IAFA and IASA scales. The misfit item substantiated the claim that reverse-worded items could lead to inferior psychometric properties of a scale (Pestle, Chorpita, & Schiffman, 2008).

After the deletion of two poorly fitting items, the psychometric properties of the four scales were re-examined and these results are presented in [Table 1](#). Rasch reliability has a similar interpretation to Cronbach's alpha, but is more conservative and less misleading (Linacre, 1997). The Rasch person/item reliabilities for the four scales were all above 0.75/0.9. The Infit/Outfit MNSQs were all within the range of 0.5–1.5. The variance explained by each scale was around 50%. The eigenvalue for the first contrast is an

Table 1. Psychometric properties of the four scales.

Scale	No. of items	Rasch person/item reliability	Infit MNSQ range	Outfit MNSQ range	Variance explained by measures	Eigenvalue for the 1 st contrast
AAFA	7	0.75/0.97	1.33/0.72	1.47/0.69	48.5%	1.5
IAFA	8	0.76/0.99	1.28/0.73	1.38/0.74	49.9%	1.6
AASA	7	0.77/0.96	1.44/0.71	1.48/0.64	53.0%	1.5
IASA	8	0.79/0.98	1.36/0.83	1.34/0.85	49.2%	2.0

indicator to measure if a secondary dimension exists and is substantial enough to distort unidimensional measurement (Raiche, 2005). The critical value for this index should be no larger than 2.0. The eigenvalues for the first contrast of four scales were within the acceptable range. The results indicated that the instrument was acceptable to measure students' attitudes to formative and summative assessment.

The function of the 6-point rating scale was also checked. As shown by the Category Probability Curves in Figure 1(a–d), threshold disordering appeared. Ideally, each curve from left to right (from response categories 1–6) should have a unique peak to indicate each option is sufficiently used. In this study, Category 2 (for all the four scales) and Category 5 (for AAFA and IASA) were subsumed by other categories, indicating that Category 2 and 5 were not observed frequently enough in this sample. Nevertheless, Linacre (1999) argued that 'threshold disordering does not indicate that the category definitions are out of sequence. Rather, the category defines a narrow section of the variable' (p. 675). Later on, the category functioning was further examined within the main sample ($N = 3,019$). As shown in Appendix 2 (Figure A2), the six categories functioned much better in the main sample. Thus, six categories were kept for the main analysis.

The finalised instrument consisted of 30 items with seven items for the AAFA scale (e.g. FA is interesting), eight items for the IAFA scale (e.g. FA can improve my learning outcome), seven items for the AASA scale (e.g. SA is interesting), and eight items for the IASA scale (e.g. SA can improve my learning outcome).

Students' general attitudes to formative and summative assessment

After examining the instrument quality via the pilot study, 3,019 participants from 10 local schools responded to the finalised instrument. In Rasch analysis, the mean of item difficulty is set to zero and higher values on the scale indicate higher item difficulty and higher student ability. Therefore, a positive student measure implies that the student has a highly positive attitude, with respect to the mean of the items; however, a negative student measure indicates an unfavourable attitude. As shown in Table 2, students in the sample, on average, hold negative affective attitudes towards FA ($M = -.34$) and SA ($M = -.22$), but positive instrumental attitudes towards FA ($M = .34$) and SA ($M = .47$). In other words, students did not like assessments no matter they serve formative or summative purpose, but they understood and appreciated that assessment is necessary and useful for learning. Also, as the correlation matrix shows in Table 2, the four scales were highly correlated ($.66 \leq r \leq .81$), which laid the basis for hierarchical regression analysis to examine the relationship of the four scales.

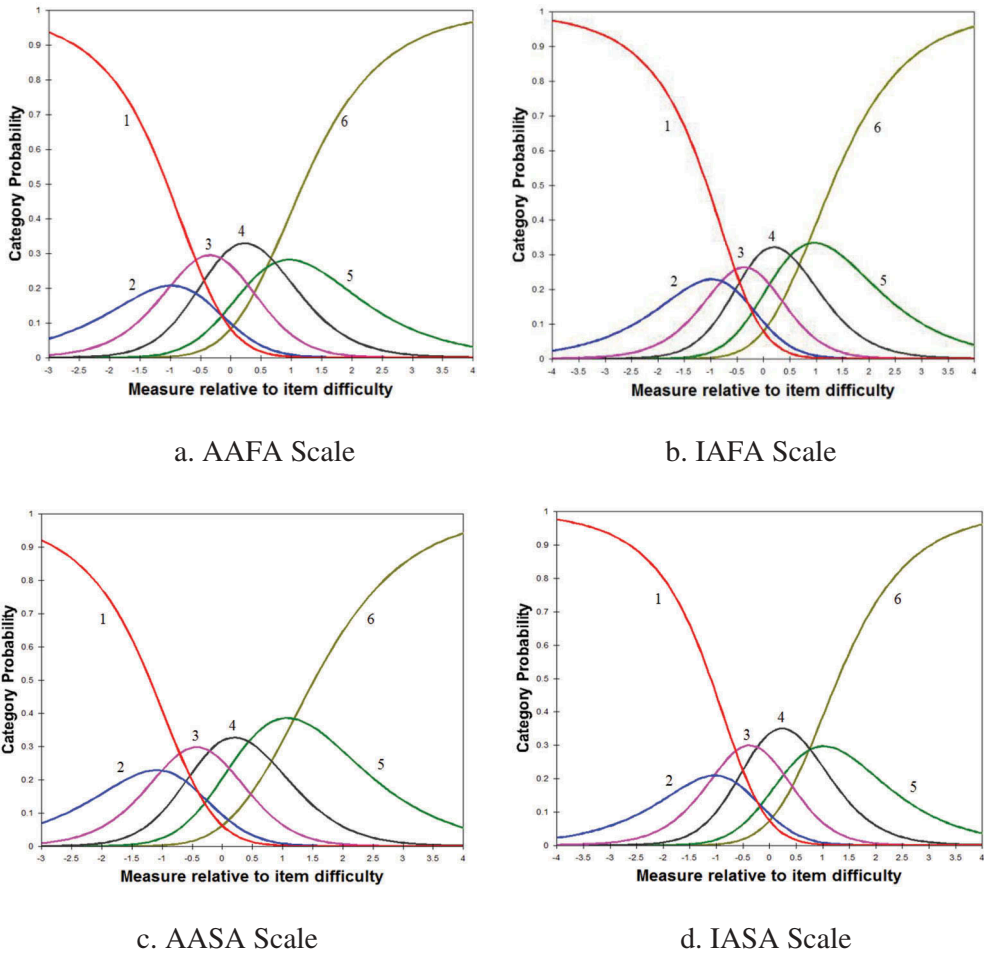


Figure 1. Category probability curves of the four scales with the pilot sample. For all four charts, the six sequential lines from left to right represent response category 1 to category 6, respectively.

Table 2. Means, standard deviations, and correlation matrix of four scales.

	<i>M</i>	<i>SD</i>	I	II	III	IV
AAFA	-.34	1.31	1.00			
IAFA	.34	1.41	.79**	1.00		
AASA	-.22	1.40	.77**	.66**	1.00	
IASA	.47	1.59	.69**	.81**	.78**	1.00

Means are in logits; ***p* < .01.

In order to more clearly understand students’ general attitudes towards both FA and SA, four variable maps (Figures 2–5) were generated using Winsteps (Linacre, 2011). The student column on the left represents the ability distribution of 3,019 students. The item column in the middle represents item difficulty distribution. The item threshold on the right represents the range of difficulty levels of response categories for each item (e.g. ‘AAFA1.1’ stands for the first threshold for item AAFA1). Students’ attitudes about

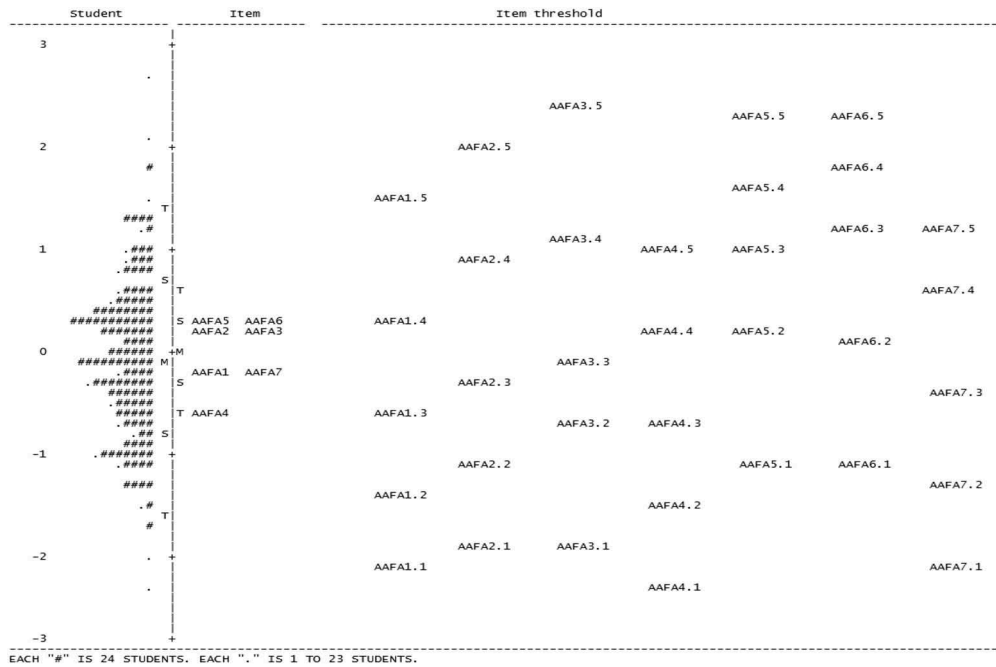


Figure 2. Item-person map of AAFA scale. AAFA1.1, for example, stands for 1st threshold for item AAFA1.

formative and summative assessments strongly echoed each other. With regard to students’ affective attitudes, Figures 2 and 4 show that items AAFA2 vs. AASA2 g(FA/SA is an enjoyable process) and items AAFA3 vs. AASA3 (FA/SA makes my study easier) were among the hardest items to endorse, which meant that students had the greatest difficulty in agreeing with this statement. Items AAFA4 and AASA7 (FA/SA deserves my full commitment), in contrast, were the easiest to endorse. That is to say, students highly agreed that they should treat assessment seriously. With regard to students’ instrumental attitudes (Figures 3 and 5), items IAFA3 and IASA2 (FA/SA enhances my learning motivation) were the hardest to endorse as students felt neither formative nor summative assessments elevated their enthusiasm for learning. Items IAFA2 and IASA1 (FA/SA helps me know my strengths and weaknesses) were highly agreed upon by students, which meant they perceived FA and SA as a useful tool to identify their learning needs.

For the student and item columns, student ability ranged from -3.0 to $+4.0$ logits in all four scales, while item difficulty ranged from -1.0 to $+0.8$ logits. It is clear that the item difficulty range was far smaller than the student ability range. However, the difficulty levels of the six response categories for each item, as presented in Figures 2–5, covered a much wider range. In other words, students’ abilities are well measured by the items with six categories, which cover a wide range on the latent trait continuum.

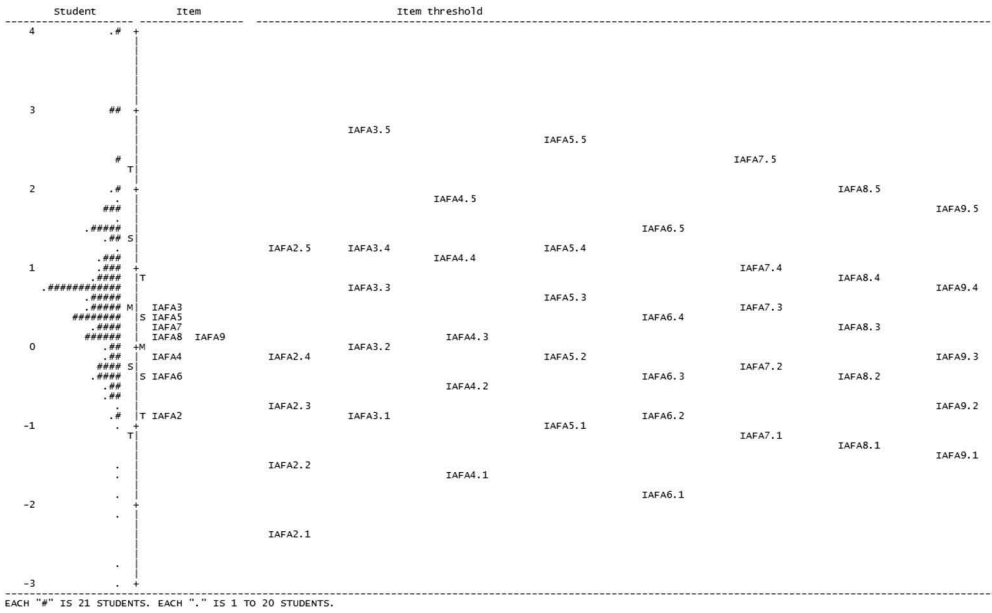


Figure 3. Item-person map of IAFA scale.
IAFA2.2, for example, stands for 2nd threshold for item IAFA2.

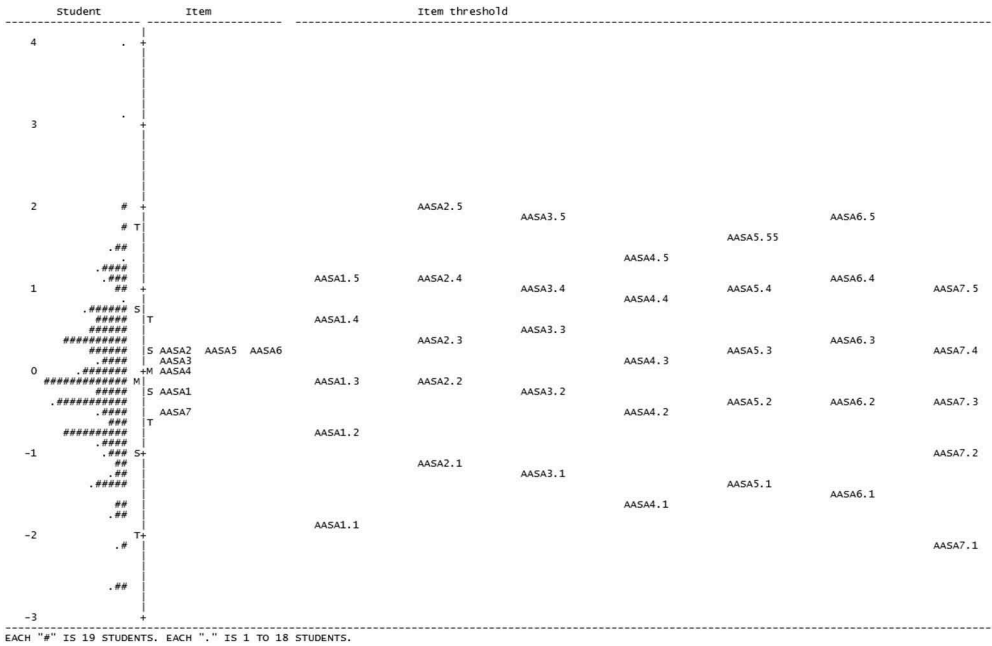


Figure 4. Item-person map of AASA scale.
AASA3.3, for example, stands for 3rd threshold for item AASA3.

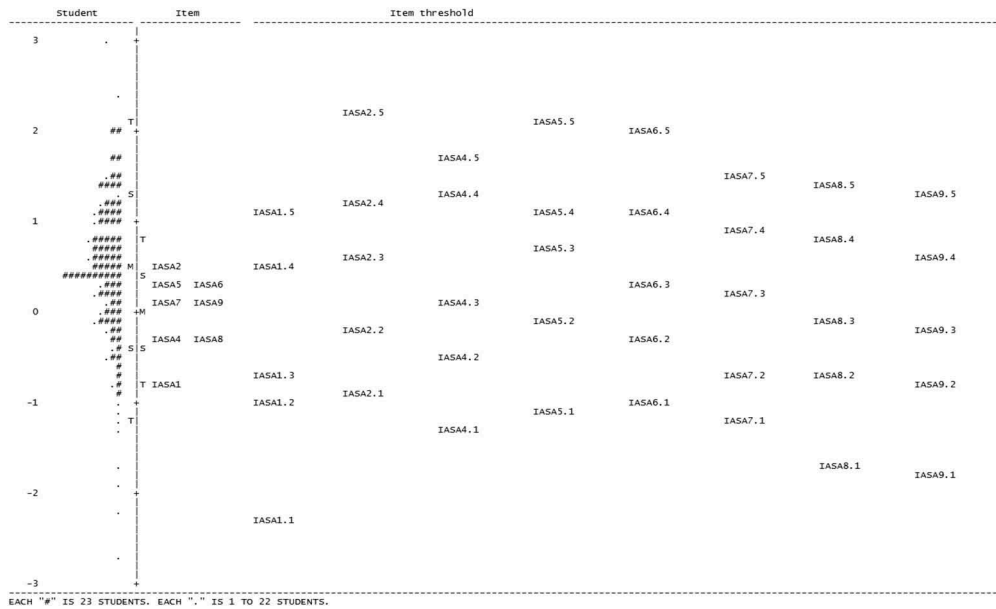


Figure 5. Item-person map of IASA scale. IASAA4.4, for example, stands for 4th threshold for item IASAA4.

Table 3. T-test of gender differences in four scales.

	Male (N = 1558)		Female (N = 1427)		t	p	Effect size (Cohen's d) ^a
	M	SD	M	SD			
AAFA	-.43	1.37	-.31	1.27	-2.52	.012*	.09
IAFA	.22	1.43	.42	1.39	-3.82	.000**	.14
AASA	-.28	1.46	-.20	1.36	-1.41	.160	.06
IASA	.35	1.62	.56	1.60	-3.44	.001**	.13

Means are in logits; *p < .05; **p < .01.

^aCohen's (1988) threshold for effect size (d): Small: .20; Medium: .50; Large: .80.

Gender differences in attitudes to FA and SA

As shown in Table 3, females (M = -.31, SD = 1.27) possessed significantly (p < .05) higher affective attitudes towards FA than males (M = -.43, SD = 1.37). Girls (M = .42, SD = 1.39) also outstripped (p < .01) boys (M = .22, SD = 1.43) in positive instrumental attitudes towards FA. Similarly, boys (M = .35, SD = 1.62) showed significantly (p < .01) less positive instrumental attitudes towards SA than did girls (M = .56, SD = 1.60). However, gender difference did not exist (p = .160) for affective attitudes towards SA and such differences in AAFA, IAFA, and IASA had trivial effect sizes (Cohen's d ≤ .2).

Grade differences in attitudes to FA and SA

Results from Table 4 demonstrate that the positive affective attitudes towards FA for Grade 4 students (M = -.24, SD = 1.33) were significantly (p < .01) higher than for

Table 4. One-way ANOVA of grade differences in 4 scales.

	Grade 4 (N = 889)		Grade 5 (N = 992)		Grade 6 (N = 1126)		F	p	Effect size (η^2) ^a
	M	SD	M	SD	M	SD			
AAFA	-.24	1.33	-.35	1.34	-.48	1.29	8.73	.000**	.01
IAFA	.43	1.47	.31	1.42	.21	1.37	5.89	.003**	.004
AASA	-.11	1.47	-.25	1.42	-.34	1.35	6.78	.001**	.004
IASA	.53	1.68	.44	1.63	.39	1.53	1.82	.163	.001

Means are in logits; * $p < .05$; ** $p < .01$.

^aCohen's (1988) threshold for effect size (η^2): Small: .01; Medium: .059; Large: .138.

Grade 6 students ($M = -.48$, $SD = 1.29$). With regard to positive instrumental attitudes towards FA, students in Grade 4 ($M = .43$, $SD = 1.47$) also exceeded ($p < .01$) their counterparts in Grade 6 ($M = .21$, $SD = 1.37$). Moreover, Grade 6 ($M = -.34$, $SD = 1.35$) students' positive affective attitudes towards SA was significantly ($p < .01$) lower than that of Grade 5 ($M = -.25$, $SD = 1.42$) as well as Grade 4 students ($M = -.11$, $SD = 1.47$). Further, students in Grade 5 ($M = -.25$, $SD = 1.42$) had significantly ($p < .01$) less positive affective attitudes towards SA than those in Grade 4 ($M = -.11$, $SD = 1.47$). However, no grade differences ($p = .16$) were found among students in terms of their instrumental attitudes towards SA. Similar to gender difference, all grade differences had trivial effect sizes ($\eta^2 \leq .01$).

Hierarchical regression analysis of attitudes to FA predicting attitudes to SA

Since the gender and grade differences were trivial in effect size, the two demographic variables were controlled (see Model 1 in Table 5) at the first level of hierarchical analysis. The results of regression Model 2 in Table 5 revealed that students' affective and instrumental attitudes to FA could significantly ($p < .01$) predict their affective attitudes towards SA. Students' attitudes to FA accounted for 60% of the variance in students' affective attitudes towards SA. This implied that if a student disliked FA and/or underestimate the value of FA, it was less likely that student would hold a positive affective attitude towards SA.

Accordingly, students' affective and instrumental attitudes to FA could significantly ($p < .01$) predict their instrumental attitudes towards SA (Table 6). Sixty-six per cent of the variance in instrumental attitudes towards SA was explained by students' attitudes to FA. That is to say, if a student accepted FA well, there is a high possibility that this student would perceive SA useful to his/her study.

Table 5. Hierarchical regression analysis of attitudes to FA predicting affective attitude towards SA.

Predictors	Model 1			Model 2		
	β	B	SE	β	B	SE
Gender	.03	.08	.05	-.01	-.03	.03
Grade	-.07**	-.12	.03	-.01	-.02	.02
AAFA				.68**	.72	.02
IAFA				.12**	.12	.02
R ²	.01**			.61**		
ΔR^2	.01**			.60**		

Dependent variable is AASA; ** $p < .01$.

Table 6. Hierarchical regression analysis of attitudes to FA predicting instrumental attitude towards SA.

Predictors	Model 1			Model 2		
	β	B	SE	β	B	SE
Gender	.06**	.21	.06	.01	.02	.04
Grade	-.03	-.07	.04	.02	.04	.02
AAFA				.14**	.17	.02
IAFA				.69**	.79	.02
R ²	.01**			.66**		
ΔR^2	.01**			.65**		

Dependent variable is IASA; ** $p < .01$.

Discussion

The current study was designed to further understand students' affective and instrumental attitudes towards FA and SA, the gender and grade differences in students' attitudes towards FA and SA, and the relationships between the two forms of assessment.

To address the four research questions, one instrument consisting of four scales (AAFA, IAFA, AASA, IASA) was firstly developed. The Rasch model analysis (Rasch, 1960) was applied to examine the psychometric properties of the instrument. As a result, a 30-item instrument with satisfactory psychometric properties was finalised and utilised for this study. This newly-developed instrument with its four psychometrically unidimensional measurement scales contributes as a valid research tool for scholars to explore questions regarding attitudes towards FA and SA.

The first research question was to understand students' attitudes towards FA and SA. On average, students had negative affective attitudes towards FA and SA, yet positive instrumental attitudes towards FA and SA. These results imply that students do not like being assessed in either formative or summative ways, but they were clear that assessments were useful for their learning. This finding was in line with previous studies on Hong Kong primary to college students (Brown & Wang, 2016; Carless & Lam, 2014; Hue et al., 2015; Lee, 2007) as well as New Zealand secondary school students (Brown & Irving et al., 2009). In conclusion, students dislike FA and SA, yet use them instrumentally. This finding echoes McMillan's (2016) summary on student's mixed attitudes toward assessment such that they value but dislike assessment. Based on this result, it is recommended that teachers attempt to ameliorate students' overall negative affective attitudes towards FA when implementing FA in classrooms. Helping students to truly understand the function of FA as assessment for learning might ease their affective resistance to FA.

The next step was to study gender differences in attitudes towards FA and SA (Research Question 2). Gender difference was found to be statistically significant but to have a trivial effect size as what had been reported in the literature (Awofala & Babajide, 2013; Dhindsa et al., 2007). Nevertheless, a salient pattern is worth reporting. Female students, on average, possessed more positive attitudes than did males, especially affective attitudes towards FA and SA, as well as instrumental attitudes towards FA. This means that girls in Hong Kong primary schools could be more likely than boys to perceive FA and SA as helpful for their learning. This finding of gender difference in assessment attitudes is consistent with several previous studies (Alkharusi, 2013; Meece et al., 2003; Yan, 2018) which also found that, female students held more positive perceptions of assessment than did males with small gender effect (less than 2% of the variance). One possible reason for this phenomenon could be related to the

gender difference in favour of girls found in some studies on self-regulated learning (Klassen et al., 2009; Virtanen & Nevgi, 2010). FA is a core strategy to facilitate self-regulated learning (Nicol & Macfarlane-Dick, 2006; Panadero & Jonsson, 2013). It might be reasonable to infer that girls tend to possess more positive attitudes towards FA than boys because FA can help them develop self-regulated learning capabilities. Another speculative reason could be girls' higher level of social compliance orientation. As argued by some scholars (Abraham, 1995; Whitelaw, Milosevic, & Daniels, 2000), girls tended to be more compliant than a boy in schooling-related issues, which implied more passive acceptance of teaching and learning requirements. This indicates that girls showed somewhat more favourable attitude toward FA and SA simply because, comparing with boys, they are more likely to habitually agree with assessment styles or requirements from teachers or schools.

Regarding the relationship between school grade and attitudes towards assessments (Research Question 3), a consistent pattern was identified. Students' preferences for FA and SA showed a consistently decreasing tendency as grades increased in spite of small effect sizes of grade difference which was also found in previous studies (Awofala & Babajide, 2013; Dhindsa et al., 2007). Students in Grade 6 showed less positive attitudes towards both FA and SA than Grade 5, who showed less positive attitudes than Grade 4. This is not an unusual observation in an exam-oriented education system (Berry, 2011; Chen, Yeh, Hwang, & Lin, 2013; Lee, 2007). The study of Chen et al. (2013) concluded that Taiwanese students' perceptions of academic pressure increased as their grades increased. Students in lower grades experienced little test pressure, while higher grade students were faced with competition and stress from entrance exams for the next stage of education. In this sense, it is understandable that Grade 4 primary school students in Hong Kong could make more use of FA and SA than their Year 6 counterparts. This suggests that the benefits of FA in enhancing students' academic motivation and performance (Cauley & McMillan, 2010), instilling passion for life-long learning (Young & Jackman, 2014), and cultivating self-regulated learners (Nicol & Macfarlane-Dick, 2006; Panadero & Jonsson, 2013) could be achieved at an early stage of schooling. Further, systematic grade differences in attitudes towards FA also reflect that the implementation of such assessments requires a low-pressure teaching and learning environment. Indeed, even in the contexts where academic pressure is relatively low and less exam-oriented, such as New Zealand, some studies (Brown & Harris, 2012; Harris, Brown, & Harnett, 2014, 2015) reported that younger students tended to perceive FA more positively than senior students who experienced more stressful judgements by the school assessment system. In this sense, if schools hope to facilitate FA among senior students, they should be given ample time and space to perform various types of FA.

Admittedly, it cannot be concluded that the gender or grade differences reported above were substantial enough because the effect sizes are trivial. However, the systematic pattern reported here might serve as a reference or a cue which warrants further professional thought and empirical research.

In terms of the predictive power of attitude towards FA for attitudes towards SA (Research Question 4), this study revealed that students' attitudes towards FA effectively predicted their attitudes towards SA. A student with a positive affective and instrumental attitude towards FA can be predicted to have a higher chance of affectively accepting SA and perceiving SA as more useful for their learning. This finding could be interpreted from two perspectives. First, the results corroborated the claim that FA and SA should be reconciled, rather than separated from each

other, in classroom practices (Carless, 2011; Taras, 2005). It is no doubt that teachers in different contexts are implementing their particular FA practices (Black & Wiliam, 2018) even though the major objective of such FA practices in Hong Kong classrooms is probably preparing students for SA (e.g. to obtain good grades in high-stakes examinations) (Leong, Ismail, Costa, & Tan, 2018). In this sense, it is not surprising to find that attitudes towards FA predicted attitudes towards SA well in this study. Second, although the regression results alone can't guarantee a causal relationship (Allison, 1999), the strong correlation between students' attitudes towards FA and SA, as demonstrated in this study, has practical implications. As the frequency of FA is much higher than SA in practice, it seems more feasible to manipulate students' attitudes towards FA. Given that high-stakes SA may maintain their important role in East Asia, including Hong Kong, in the foreseeable future (Leong et al., 2018), the results of this study shed lights on the possible way to predict and change students' attitudes towards SA. When teachers and schools have a better understanding about the correlation between students' attitudes towards FA and SA, they can design appropriate pedagogical interventions to alter students' attitudes towards FA and accordingly shape students' attitudes towards SA.

Interpretation of the results should be made with caution due to the limitations of this study. Firstly, the research instrument exclusively relied on primary school students' self-reporting. The results may thus be vulnerable to the common-method variance effect. Efforts have been made to tackle this issue. However, taking primary students' language literacy and cognition capacity into account, self-reports could be the most accurate way to approach studies about psychological characteristics such as attitude, given that individuals should have better insights into their own beliefs than outside observers (Lee, Chang, Lin, & Cheng, 2014). Secondly, the generalisation of the conclusions drawn from this study should be limited to primary students in Hong Kong or any jurisdictions where an exam-oriented educational culture is dominant.

One direction for future research could be further examination of the psychometric properties of the instrument with other groups of population in order to widen its application. This study only employed primary school students to test the appropriateness of the instrument. Since the newly developed instrument enables researchers to study FA and SA in a parallel manner, it could be tested in future research with a wider population, such as secondary or college students. Moreover, factors that influence students' attitudes to FA and SA could be explored so that students' perception development process could be revealed. The current study examined two demographic variables which could be potential traits causing different attitudes to FA and SA. Relevant elements that could affect students' assessment attitudes still await discovery. Lastly, this study only begins to scrutinise the association between FA and SA. Future studies, especially ones with a controlled experimental design, could aim to uncover the mechanism of how the two forms of assessment influence and interact with each other. How could students' attitudes to FA be transferred to attitudes towards SA? Is there any mediator or moderation that could play a role in the interaction between FA and SA? Answers to these questions will have significant implications for teaching and learning.

Conclusion

This study contributes to the understanding of primary school students' attitudes towards FA and SA, especially in the Hong Kong context. A new attitude instrument was constructed based on the Theory of Planned Behaviour, which showed that no significant gender and grade differences towards FA and SA existed among local students, although there seemed to be a systematic pattern in favour of female and junior learners. Students' affective and instrumental attitudes to FA predicted their affective and instrumental attitudes to SA. These findings contribute to the literature concerning the relationship between FA and SA and also have practical implications for teachers' instruction in classrooms.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Wu Yuan Guo is a PhD candidate at The Education University of Hong Kong. Her research interests focus on educational assessment and educational psychology.

Zi Yan is an associate professor and associate head in the Department of Curriculum and Instruction at The Education University of Hong Kong. His main publication and research interests are in Rasch measurement and assessment in the school and higher education context with an emphasis on student self-assessment and self-regulated learning.

ORCID

Wu Yuan Guo  <http://orcid.org/0000-0001-8022-1433>

Zi Yan  <http://orcid.org/0000-0001-9305-884X>

References

- Abraham, J. (1995). *Divide and school: Gender and class dynamics in comprehensive education*. London: Falmer Press.
- Ajzen, I. (1985). From intentions to actions: A theory of planned behavior. In J. Kuhl & J. Beckmann (Eds.), *Action control: From cognition to behavior* (pp. 11–39). Berlin: Springer Berlin Heidelberg.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211.
- Alkharusi, H. (2013). Canonical correlational models of students' perceptions of assessment tasks, motivational orientations, and learning strategies. *Online Submission*, 6(1), 21–38.
- Allison, P. D. (1999). *Multiple regression: A primer*. Thousand Oaks, CA: Pine Forge Press.
- Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behavior: A meta-analytic review. *British Journal of Social Psychology*, 40, 471–499.
- Awofala, A. O., & Babajide, V. F. (2013). Examining attitude towards continuous assessment practices among Nigerian preservice STM teachers. *Journal of Education and Practice*, 4(13), 37–49.
- Bacon, D. R., & Bean, B. (2006). GPA in research studies: An invaluable but neglected opportunity. *Journal of Marketing Education*, 28, 35–42.

- Berry, R. (2011). Assessment trends in Hong Kong: Seeking to establish formative assessment in an examination culture. *Assessment in Education: Principles, Policy & Practice*, 18(2), 199–211.
- Biggs, J. (1998). Assessment and classroom learning: A role for summative assessment? *Assessment in Education: Principles, Policy & Practice*, 5(1), 103–110.
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551–575.
- Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice*, 5(1), 7–74.
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). New York, NY: Routledge.
- Brady, L., & Kennedy, K. (2005). *Celebrating student achievement: Assessment and reporting* (2nd ed.). Frenchs Forest, N.S.W.: Pearson Education Australia.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2001). *How people learn: Brain, mind, experience, and school*. Washington, D.C.: National Academy Press.
- Brown, G. T., & Harris, L. R. (Eds.). (2016). *Handbook of human and social conditions in assessment*. New York: Routledge.
- Brown, G. T., & Hirschfeld, G. H. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education Principles Policy & Practice*, 15(1), 3–17.
- Brown, G. T., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. (2009). Use of interactive–Informal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning and Instruction*, 19(2), 97–111.
- Brown, G. T., Kennedy, K. J., Fok, P. K., Chan, J. K. S., & Yu, W. M. (2009). Assessment for student improvement: Understanding Hong Kong teachers' conceptions and practices of assessment. *Assessment in Education: Principles, Policy & Practice*, 16(3), 347–363.
- 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.
- Brown, G. T. L., & Harris, L. R. (2012). Student conceptions of assessment by level of schooling: Further evidence for ecological rationality in belief systems. *Australian Journal of Educational and Developmental Psychology*, 12, 46–59.
- Brown, G. T. L., & Wang, Z. (2016). Understanding Chinese university student conceptions of assessment: Cultural similarities and jurisdictional differences between Hong Kong and China. *Social Psychology of Education*, 19(1), 151–173.
- Carless, D. (2011). *From testing to productive student learning: Implementing formative assessment in Confucian-heritage settings*. New York: Routledge.
- Carless, D., & Lam, R. (2014). The examined life: Perspectives of lower primary school students in Hong Kong. *Education*, 42(3), 313–329.
- Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House*, 83(1), 1–6.
- Chen, S. K., Yeh, Y. C., Hwang, F. M., & Lin, S. S. (2013). The relationship between academic self-concept and achievement: A multi-cohort and multi-occasion study. *Learning and Individual Differences*, 23, 172–178.
- Choi, C. C. (1999). Public examinations in Hong Kong. *Assessment in Education: Principles, Policy & Practice*, 6(3), 405–417.
- Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, 24(2), 205–249.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Routledge Academic.
- Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). New York, NY: Routledge.
- Conlon, T. (2006). Formative assessment of classroom concept maps: The reasonable fallible analyser. *Journal of Interactive Learning Research*, 17(1), 15–36.
- Curriculum Development Council, Hong Kong. (2001). *Physical education curriculum and assessment guide (Secondary 4-6)*. Hong Kong: Education Department.

- Davies, J., & Brember, I. (1998). National curriculum testing and self-esteem in Year 2—The first 5 years: A cross-sectional study. *Educational Psychology, 18*(4), 365–375.
- Deb, S., Strodl, E., & Sun, J. (2014). Academic-related stress among private secondary school students in India. *Asian Education and Development Studies, 3*(2), 118–134.
- Deneen, C., Brown, G. T., Bond, T. G., & Shroff, R. (2013). Understanding outcome-based education changes in teacher education: Evaluation of a new instrument with preliminary findings. *Asia-Pacific Journal of Teacher Education, 41*(4), 441–456.
- Dhindsa, H. S., Omar, K., & Waldrip, B. (2007). Upper secondary Bruneian science students' perceptions of assessment. *International Journal of Science Education, 29*(10), 1261–1280.
- Eady, S. (1999). An investigation of possible correlation of general anxiety with performance in eleven-plus scores in year 6 primary school pupils. *Educational Psychology, 19*(3), 347–359.
- Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading: MA: Addison-Wesley.
- Glover, T. A., Reddy, L. A., Kettler, R. J., Kurz, A., & Lekwa, A. J. (2016). Improving high-stakes decisions via formative assessment, professional development, and comprehensive educator evaluation: The school system improvement project. *Teachers College Record, 118*(14), 1–26.
- Gregor, A. (2005). Examination anxiety: Live with it, control it or make it work for you? *School Psychology International, 26*(5), 617–635.
- Harris, L. R., Brown, G. T. L., & Harnett, J. A. (2014). Understanding classroom feedback practices: A study of New Zealand student experiences, perceptions, and emotional responses. *Educational Assessment, Evaluation and Accountability, 26*(2), 107–133.
- Harris, L. R., Brown, G. T. L., & Harnett, J. A. (2015). Analysis of New Zealand primary and secondary student peer- and self-assessment comments: Applying Hattie and Timperley's feedback model. *Assessment in Education: Principles, Policy & Practice, 22*(2), 265–281.
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan, 89*(2), 1–8.
- Hue, M. T., Leung, C. H., & Kennedy, K. J. (2015). Student perception of assessment practices: Towards 'no loser' classrooms for all students in the ethnic minority schools in Hong Kong. *Educational Assessment, Evaluation and Accountability, 27*(3), 253–273.
- Hui, S. K. F., Brown, G. T. L., & Chan, S. W. M. (2017). Assessment for learning and for accountability in classrooms: The experience of four Hong Kong primary school curriculum leaders. *Asia Pacific Education Review, 18*(1), 41–51.
- Jackson, J. (2013). Adjusting to differing cultures of learning: The experience of semester-long exchange students from Hong Kong. In L. Jin & M. Cortazzi (Eds.), *Researching intercultural learning: Investigations in language and education* (pp. 235–252). Basingstoke, Hampshire: Palgrave Macmillan.
- Klassen, R. M., Ang, R. P., Chong, W. H., Krawchuk, L. L., Huan, V. S., Wong, I. Y., & Yeo, L. S. (2009). A cross-cultural study of adolescent procrastination. *Journal of Research on Adolescence, 19*(4), 799–811.
- Knight, P. T. (2002). Summative assessment in higher education: Practices in disarray. *Studies in Higher Education, 27*(3), 275–286.
- Kwon, S. K., Lee, M., & Shin, D. (2017). Educational assessment in the Republic of Korea: Lights and shadows of high-stake exam-based education system. *Assessment in Education: Principles, Policy & Practice, 24*(1), 60–77.
- Law, W. W. (2007). Legislation and educational change: The struggle for social justice and quality in China's compulsory schooling. *Education and Law, 19*(3–4), 177–199.
- Lee, I. (2007). Feedback in Hong Kong secondary writing classrooms: Assessment for learning or assessment of learning? *Assessing Writing, 12*(3), 180–198.
- Lee, Y. K., Chang, C. T., Lin, Y., & Cheng, Z. H. (2014). The dark side of smartphone usage: Psychological traits, compulsive behavior and technostress. *Computers in Human Behavior, 31*, 373–383.
- Leong, W. S., Ismail, H., Costa, J. S., & Tan, H. B. (2018). Assessment for learning research in East Asian countries. *Studies in Educational Evaluation, 59*, 270–277.

- Linacre, J. M. (1997). KR-20/Cronbach alpha or Rasch person reliability: Which tells the “truth”? *Rasch Measurement Transactions*, 11(3), 580–581.
- Linacre, J. M. (1999). Category disordering (disordered categories) vs. threshold disordering (disordered thresholds). *Rasch Measurement Transactions*, 13(1), 675–677.
- Linacre, J. M. (2011). *Winsteps® (Version 3.7)*. [Computer Software]. Beaverton, Oregon: Winsteps.com. Retrieved from <http://www.winsteps.com/>
- Marinho, P., Leite, C., & Fernandes, P. (2017). Mathematics summative assessment practices in schools at opposite ends of performance rankings in Portugal. *Research in Mathematics Education*, 19(2), 184–198.
- 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.
- Meece, J. L., Herman, P., & McCombs, B. L. (2003). Relations of learner-centered teaching practices to adolescents’ achievement goals. *International Journal of Educational Research*, 39(4–5), 457–475.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2015). *Introduction to linear regression analysis*. New York: John Wiley & Sons.
- Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218.
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144.
- Pat-EI, R. J., Tillema, H., Segers, M., & Vedder, P. (2015). Multilevel predictors of differing perceptions of assessment for learning practices between teachers and students. *Assessment in Education: Principles, Policy & Practice*, 22(2), 282–298.
- Pestle, S., Chorpita, B., & Schiffman, J. (2008). Psychometric properties of the Penn State Worry Questionnaire for children in a large clinical sample. *Journal of Clinical Child & Adolescent Psychology*, 37, 465–471.
- Raïche, G. (2005). Critical eigenvalue sizes (variances) in standardized residual principal components analysis (PCA). *Rasch Measurement Transactions*, 19(1), 10–12.
- Rasch, G. (1960). *Probabilistic models for some intelligence and achievement test*. Copenhagen, Denmark: Danish Institute for Educational Research. Expanded ed. (1980). Chicago, IL: The University of Chicago Press.
- Reay, D., & Wiliam, D. (1999). ‘I’ll be a nothing’: Structure, agency and the construction of identity through assessment. *British Educational Research Journal*, 25(3), 343–354.
- Roos, B., & Hamilton, D. (2005). Formative assessment: A cybernetic viewpoint. *Assessment in Education: Principles, Policy & Practice*, 12(1), 7–20.
- Rust, C. (2000). An opinion piece: A possible student-centred assessment solution to some of the current problems of modular degree programmes. *Active Learning in Higher Education*, 1((2)), 126–131.
- Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice*, 5(1), 77–84.
- Taras, M. (2005). Assessment–Summative and formative–Some theoretical reflections. *British Journal of Educational Studies*, 53(4), 466–478.
- Virtanen, P., & Nevgi, A. (2010). Disciplinary and gender differences among higher education students in self-regulated learning strategies. *Educational Psychology*, 30(3), 323–347.
- Volante, L., & Beckett, D. (2011). Formative assessment and the contemporary classroom: Synergies and tensions between research and practice. *Canadian Journal of Education*, 34, 239–255.
- Whitelaw, S., Milosevic, L., & Daniels, S. (2000). Gender, behaviour and achievement: A preliminary study of pupil perceptions and attitudes. *Gender and Education*, 12(1), 87–113.
- Wright, B. D. (1997). A history of social science measurement. *Educational Measurement: Issues and Practice*, 16(4), 33–45.

- Yan, Z. (2014). Predicting teachers' intentions to implement school-based assessment using the Theory of planned behaviour. *Educational Research and Evaluation, 20*(2), 83–97.
- Yan, Z. (2016). The self-assessment practices of Hong Kong secondary students: Findings with a new instrument. *Journal of Applied Measurement, 17*(3), 335–353.
- Yan, Z. (2017). How teachers' beliefs and demographic variables impact on self-regulated learning instruction. *Educational Studies*. doi:[10.1080/03055698.2017.1382331](https://doi.org/10.1080/03055698.2017.1382331)
- Yan, Z. (2018). Student self-assessment practices: The role of gender, school level and goal orientation. *Assessment in Education: Principles, Policy & Practice, 25*(2), 183–199.
- Yan, Z., & Cheng, E. C. K. (2015). Primary teachers' attitudes, intentions and practices regarding formative assessment. *Teaching and Teacher Education, 45*, 128–136.
- Young, J. E., & Jackman, M. G. A. (2014). Formative assessment in the Grenadian lower secondary school: Teachers' perceptions, attitudes and practices. *Assessment in Education: Principles, Policy & Practice, 21*(4), 398–411.



Appendices

Appendix 1.

Attitudes towards formative and summative assessment scale.

平時測驗態度量表 Attitudes towards Formative Assessment Scale	完全不同意 Totally disagree	不同意 Disagree	略不同意 Slightly disagree	略同意 Slightly agree	同意 Agree	完全同意 Totally agree
AAFA1 平時測驗令學習氣氛變得更好。 FA improves the learning atmosphere.	1	2	3	4	5	6
AAFA2 平時測驗是一個令人開心的過程。 FA is an enjoyable process.	1	2	3	4	5	6
AAFA3 平時測驗令學習變得更輕鬆。 FA makes my study easier.	1	2	3	4	5	6
AAFA4 平時測驗值得我用心地投入。 FA deserves my full commitment.	1	2	3	4	5	6
AAFA5 我喜歡平時測驗。 I like FA.	1	2	3	4	5	6
AAFA6 平時測驗是有趣的事。 FA is interesting.	1	2	3	4	5	6
AAFA7 平時測驗令我與同學之間互相幫助。 FA enables students to help each other.	1	2	3	4	5	6
IAFA1 平時測驗增加了我的學習負擔。(刪除) FA is a burden for my learning. (deleted)	1	2	3	4	5	6
IAFA2 平時測驗讓了解自己的強項和弱點。 FA helps me know my strengths and weaknesses.	1	2	3	4	5	6
IAFA3 平時測驗能提高我的學習興趣。 FA can increase my learning motivation.	1	2	3	4	5	6
IAFA4 平時測驗能準確地評核我的表現。 FA can accurately assess my performance.	1	2	3	4	5	6
IAFA5 平時測驗能鼓勵我自主學習。 FA can encourage me to study on my own.	1	2	3	4	5	6
IAFA6 平時測驗能公平地評核我的表現。 FA can fairly assess my performance.	1	2	3	4	5	6
IAFA7 平時測驗能提高我在學習上的自信心。FA can enhance my confidence in learning.	1	2	3	4	5	6
IAFA8 平時測驗能提高學習效果。 FA can improve my learning outcome.	1	2	3	4	5	6

(Continued)

(Continued).

平時測驗態度量表 Attitudes towards Formative Assessment Scale		完全不同意 Totally disagree	不同意 Disagree	略不同意 Slightly disagree	略同意 Slightly agree	同意 Agree	完全同意 Totally agree
平時測驗讓我更努力地學習。 FA makes me work harder.		1	2	3	4	5	6
期終測驗態度量表 Attitudes to Summative Assessment Scale							
AASA1	期終測驗令我與同學之間互相幫助。 SA enables students to help each other.	1	2	3	4	5	6
AASA2	期終測驗是一個令人開心的過程。 SA is an enjoyable process.	1	2	3	4	5	6
AASA3	期終測驗令學習變得更輕鬆。 SA makes my study easier.	1	2	3	4	5	6
AASA4	期終測驗令學習氣氛變得更好。 SA improves the learning atmosphere.	1	2	3	4	5	6
AASA5	我喜歡期終測驗。 I like SA.	1	2	3	4	5	6
AASA6	期終測驗是有趣的事。 SA is interesting.	1	2	3	4	5	6
AASA7	期終測驗值得我用心地投入。 SA deserves my full commitment.	1	2	3	4	5	6
IASA1	期終測驗讓我了解自己的強項和弱點。 SA helps me know my strengths and weaknesses.	1	2	3	4	5	6
IASA2	期終測驗能提高我的學習興趣。 SA can increase my learning motivation.	1	2	3	4	5	6
IASA3	期終測驗增加了我的學習負擔。(刪除) SA is a burden for my learning. (deleted)	1	2	3	4	5	6
IASA4	期終測驗能準確地評核我的表現。 SA can accurately assess my performance.	1	2	3	4	5	6
IASA5	期終測驗能鼓勵我自主學習。 SA can encourage me to study on my own.	1	2	3	4	5	6
IASA6	期終測驗能提高我在學習上的自信心。 SA can enhance my confidence in learning.	1	2	3	4	5	6
IASA7	期終測驗能提高學習效果。 SA can improve my learning outcome.	1	2	3	4	5	6
IASA8	期終測驗能公平地評核我的表現。 SA can fairly assess my performance.	1	2	3	4	5	6

(Continued)



(Continued).

期終測驗態度量表

Attitudes to Summative Assessment Scale

IASA9

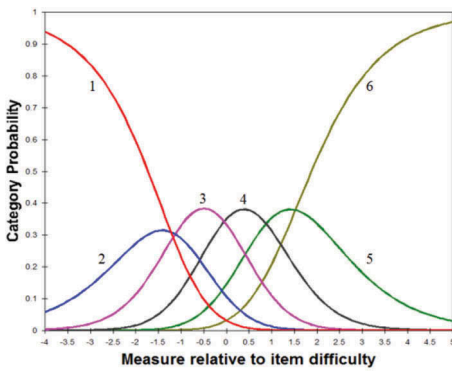
期終測驗讓我更努力地學習。

SA makes me work harder.

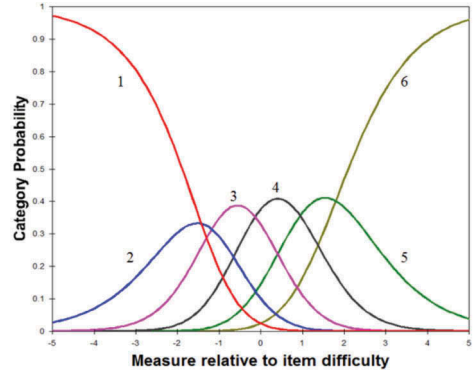
	完全 不同意 Totally disagree	不同意 Disagree	略不 同意 Slightly disagree	略同意 Slightly agree	同意 Agree	完全 同意 Totally agree
	1	2	3	4	5	6

Appendix 2.

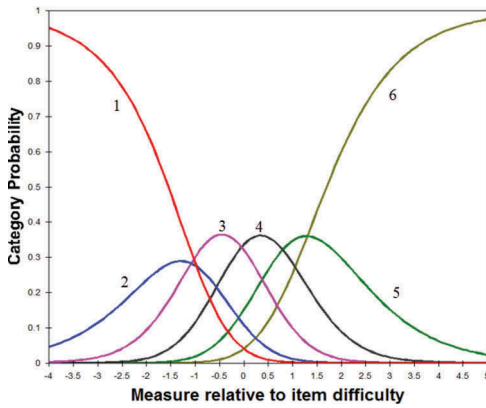
Category Probability Curves of the four scales with the main sample



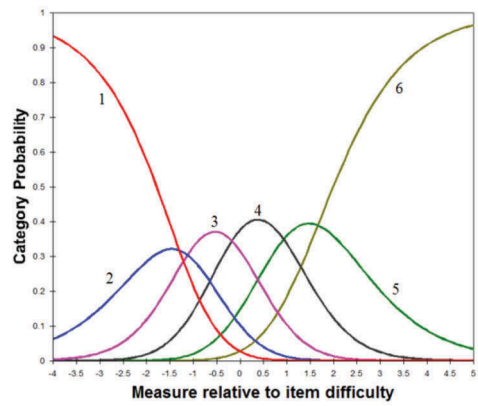
a. AAFA Scale



b. IAFA Scale



c. AASA Scale



d. IASA Scale

Figure A2. Category probability curves of the four scales with the main sample.

For all four charts, the six sequential lines from left to right represent response category 1 to category 6, respectively.