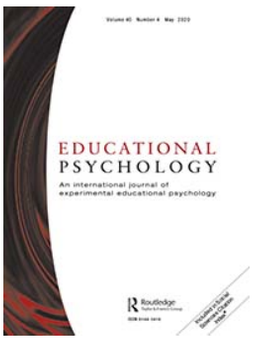


Routledge
Taylor & Francis Group

Educational Psychology

An International Journal of Experimental Educational Psychology

ISSN: 0144-3410 (Print) 1469-5820 (Online) Journal homepage: https://www.tandfonline.com/loi/cedp20

Student self-assessment: why do they do it?

Zi Yan, Gavin T. L. Brown, John Chi-Kin Lee & Xue-Lan Qiu

To cite this article: Zi Yan, Gavin T. L. Brown, John Chi-Kin Lee & Xue-Lan Qiu (2020) Student self-assessment: why do they do it?, Educational Psychology, 40:4, 509-532, DOI: 10.1080/01443410.2019.1672038

To link to this article: https://doi.org/10.1080/01443410.2019.1672038

EDUCATIONAL PSYCHOLOGY
2020, VOL. 40, NO. 4, 509–532
https://doi.org/10.1080/01443410.2019.1672038

Routledge
Taylor & Francis Group

Student self-assessment: why do they do it?

Zi Yan[a], Gavin T. L. Brown[b], John Chi-Kin Lee[a] and Xue-Lan Qiu[c]

[a]Department of Curriculum and Instruction, The Education University of Hong Kong, Hong Kong, P. R. China; [b]Faculty of Education and Social Work, The University of Auckland, Auckland, New Zealand; [c]Faculty of Education, The University of Hong Kong, Hong Kong, P. R. China

ABSTRACT

This study explores factors that predict students' self-assessment intentions and practices using a framework based on the Theory of Planned Behaviour (TPB). A total of 1425 Hong Kong students (Primary 4 to Secondary 3) participated in this study. Students' intentions and practices pertaining to self-assessment and the predictors of their intentions and practices were assessed with 11 self-report scales aligned to the TPB constructs. The psychometric properties of scales were examined with Rasch analysis and the relations among the variables were investigated with path analysis based on Rasch-calibrated person measures. The results showed that attitude, subjective norms, self-efficacy, and perceived controllability were statistically significant predictors on intention to self-assess, while self-efficacy and intention had significant influence on self-assessment practice. Psychological safety was also found to have relatively weak but significant impact on both self-assessment intention and practice. This study lays a foundation for future investigations on how to promote meaningful self-assessment behaviour which is crucial for self-regulated and life-long learning.

ARTICLE HISTORY
Received 26 February 2019
Accepted 20 September 2019

KEYWORDS
Self-assessment; predictor; theory of planned behaviour; Rasch measurement

Introduction

Student self-assessment is one of the core skills of self-regulated learning and life-long learning (Panadero & Alonso-Tapia, 2013; Yan, 2019). Self-assessing one's own learning process and outcomes can increase metacognition, i.e. awareness of one's own learning or thinking process (Siegesmund, 2016). From a pedagogical perspective, effective learning can only occur when students have a realistic sense of their own performance so that they can direct their further learning on the critical aspects of their learning needs (Baas, Castelijns, Vermeulen, Martens, & Segers, 2015; Boud, Lawson, & Thompson, 2013). Numerous empirical studies have demonstrated that self-assessment has positive effects on academic performance (Brown & Harris, 2013; Ibabe & Jauregizar, 2010; Panadero, Alonso-Tapia, & Antonio, 2012), learning

CONTACT Zi Yan zyan@eduhk.hk Department of Curriculum and Instruction, The Education University of Hong Kong, 10 Lo Ping Road, Tai Po, New Territories, Hong Kong, P. R. China

autonomy (Andrade & Du, 2007; Cassidy, 2007), commitment and engagement with learning (Brown & Harris, 2013), and academic self-efficacy (Kissling & O'Donnell, 2015).

In practice, self-assessment is used for both formative and summative purposes (Panadero, Brown, & Strijbos, 2016). However, it appears more promising to keep self-assessment formative in enacting its merits in supporting student learning due to the substantial concerns about the accuracy of self-assessment for summative purposes (Brown, Andrade, & Chen, 2015; Yan & Brown, 2017). Compared with other teacher-directed formative assessment methods that are largely hindered by big class sizes and increasing teaching workloads, self-assessment has advantages because it is less likely to be influenced by such practical and external constraints. As long as students have intentions and appropriate skills, self-assessment could be undertaken even in big classes without excessively increasing teaching loads. This phenomenon could be described with the famous Chinese saying '授人以魚不如授人以漁 (*Teaching how to fish* is better than *giving a fish*)'. Giving a fish (giving feedback through teacher-directed formative assessment) can only solve problems at hand, but teaching students how to fish (i.e. equipping them with the will and skill of self-assessment) is potentially a powerful long-term and sustainable solution. In this sense, self-assessment should be regarded as a learning strategy rather than as a substitute for other types of evaluative or summative assessment (Boud, 1999; Brown & Harris, 2014; Panadero & Alonso-Tapia, 2013; Yan, 2019). Therefore, self-assessment is appealing as a lever in fulfilling an important aim of current international education reforms (i.e. the development of self-regulated and lifelong learners).

Despite the important role of self-assessment in education, there is relatively little research focusing on students' intentions to conduct self-assessment and factors that influence students' intentions and practices pertaining to self-assessment. This situation may have arisen because student self-assessment is normally teacher-directed. That is, when and how to conduct self-assessment are largely determined by teachers rather students themselves. However, if students are expected to be agentic and autonomous, they ought to be carrying out self-assessment for their own causes and under their own control. The lack of understanding of the potential facilitators and impediments of self-assessment might hinder the benefits of self-assessment being enacted and maximised. The present study aimed to fill this research gap by exploring factors that predict students' intentions and practices pertaining to self-assessment in an extended framework of the Theory of Planned Behaviour (TPB).

Self-assessment process and actions

Different conceptualisations of self-assessment exist in literature. Panadero et al. (2016) summarised the major typologies of self-assessment and concluded that there was no consensus as to what is the 'standard self-assessment'. Yan (2016, 2018b) categorised the conceptualisations of self-assessment into three major groups: (1) self-assessment is understood as a personal ability for evaluating one's own knowledge, skills or

performance; (2) self-assessment is regarded as an alternative assessment serving summative purposes; and (3) self-assessment is used as a learning strategy or process aiming for promoting productive learning. From a pedagogical perspective, self-assessment should be regarded as a process that involves identifying the characteristics of one's own work and determining the value or merit of that work (Brown & Harris, 2013); a process that can be seen as evaluative judgment (Tai, Ajjawi, Boud, Dawson, & Panadero, 2018).

Some researchers have attempted to operationalise self-assessment as a process (e.g. Fastré, van der Klink, Sluijsmans, & van Merriënboer, 2012; McMillan & Hearn, 2008; Sargeant et al., 2010), but insights into the 'black box' of inner processes in self-assessment are still rare. A recent update in this field is the cyclical process model of self-assessment proposed by Yan and Brown (2017). In their model, when students self-assess they first determine what appropriate assessment criteria apply to the learning phenomenon. After that, students will seek feedback about the quality of their performance, if they cannot readily compose a self-assessment with the knowledge and reasons that they have to hand. The sources of feedback can be external (i.e. from past tests, textbooks, comments from teacher, peers, or parents) and/or from internal resources (e.g. emotions, physical sensations, and internal states). External feedback could be obtained by inquiry and monitoring approaches. Students can ask for (i.e. inquire) feedback directly from relevant people (e.g. teachers, peers, parents) or interact with (i.e. monitoring) learning evidence (e.g. past tests, reference books) for feedback sources. With the support of feedback, students then reflect on the quality of their learning processes and products. This allows them to identify the strengths and weaknesses of their work. Pedagogically, such self-assessment is done for the purpose of improvement, rather than as a terminal or summative evaluation. These processes lead to an initial self-assessment judgment which is then subjected to continuous calibration as stakeholders react to the reasons given in light of the work products or processes they can observe. Yan and Brown's (2017) model explicitly identified the concrete actions students commonly conduct and how they do so within self-assessment. Therefore, this model makes it possible to investigate the inner processes of students' self-assessment.

Predictors of self-assessment practice

Self-assessment is, by its nature, a personal endeavour that requires volitional effort and commitment. Therefore, it is important to identify the factors which might influence or predict self-assessment. Previous reviews (e.g. Brown & Harris, 2013) have identified some aspects that facilitate self-assessment; for example, clear assessment criteria, training, and practice in self-assessment. However, very limited information is available about the psychological attributes (e.g. attitudinal or affective factors) and the relationships of these predictors within self-assessment.

As self-assessment is a volitional behaviour over which students have control, the Theory of Planned Behaviour (TPB) (Ajzen, 1991) from the field of social psychology appears an appropriate theoretical framework to investigate its predictors. TPB outlines relationships among five components including attitudes, subjective norms,

perceived behavioural control, intentionality, and behaviours around a phenomenon. An individual's intention to enact a particular behaviour is determined by three intercorrelated factors: (1) attitudes (the cognitive evaluation and overall positive or negative evaluation toward a behaviour), (2) subjective norms (the social norms that influence an individual toward or away from a behaviour), and (3) perceived behavioural control (the extent to which the individual perceives his/her ability to control the implementation of a behaviour). In general, those with favourable attitudes, positive subjective norms, and high levels of perceived behavioural control will be more likely to exhibit an intention to perform that behaviour. Once an intention is formed, it, along with the degree of perceived behavioural control will directly influence the individual's actual behaviour.

TPB has been successfully applied to diverse behaviours in western settings (Armitage & Conner, 2001; Burrus & Moore, 2016; Hagger, Sultan, Hardcastle, & Chatzisarantis, 2015), as well as in the Hong Kong context (Mok & Lee, 2013; Yan & Cheng, 2015; Yan & Sin, 2014, 2015). Although TPB appears to be a generally appropriate framework, its comprehensiveness has been questioned because some researchers report that TPB provides a better explanation of intention than behaviour (Stanec, 2009; Yan & Cheng, 2015; Yan & Sin, 2014). Sometimes there are discrepancies between intention and actual behaviour (Fife-Schaw, Sheeran, & Norman, 2007). Ajzen (1991) admitted that the TPB framework remains open for inclusion of additional factors that may contribute to a better explanation of particular intentions and behaviours. For example, Yan and Cheng (2015) reported that the TPB-based model explained teachers' formative assessment intentions well, but not their practices. The additional elements included in the TPB framework to explain behaviour under investigation have varied in different studies, including desire (Perugini & Bagozzi, 2001) and moral norms and anticipated affect (Rivis, Sheeran, & Armitage, 2009). Usually, the additional factors have close relationships with the target behaviour but were not part of the original TPB framework.

Empirical studies have demonstrated that psychological safety (i.e. a belief that it is safe to take the interpersonal risk in a group of people) had a significant impact on intention and practice pertaining to learning behaviour (Edmondson, 1999; Van Gennip, Segers, & Tillema, 2009, 2010). For example, psychological safety was found to significantly influence university students' intentions to share knowledge in virtual communities (Zhan, Fang, Wei, & Chen, 2010) and learning behaviours undertaken by improvement project teams in hospital intensive care units (Tucker, Nembhard, & Edmondson, 2007). Van Gennip et al. (2010) reported that psychological safety had positive impact on the learning effect of peer assessment since people with high psychological safety are likely to interpret differences in viewpoints as opportunities to frame a problem rather than disagreements.

In classroom learning contexts, self-assessments are often exposed to classmates (e.g. traffic lights; Harris & Brown, 2018) or to the teacher (Cowie, 2009). It is possible that students may provide depressed self-assessment judgments to avoid being labeled as egotistical (Brooks, 2002) or they may give elevated self-assessments for fear of being shamed (Harris & Brown, 2013). In such a case, psychological safety matters. Students need to feel that it is safe in terms of interpersonal relations with peers or the teacher to indicate the truth of their own work (e.g. *I don't understand*

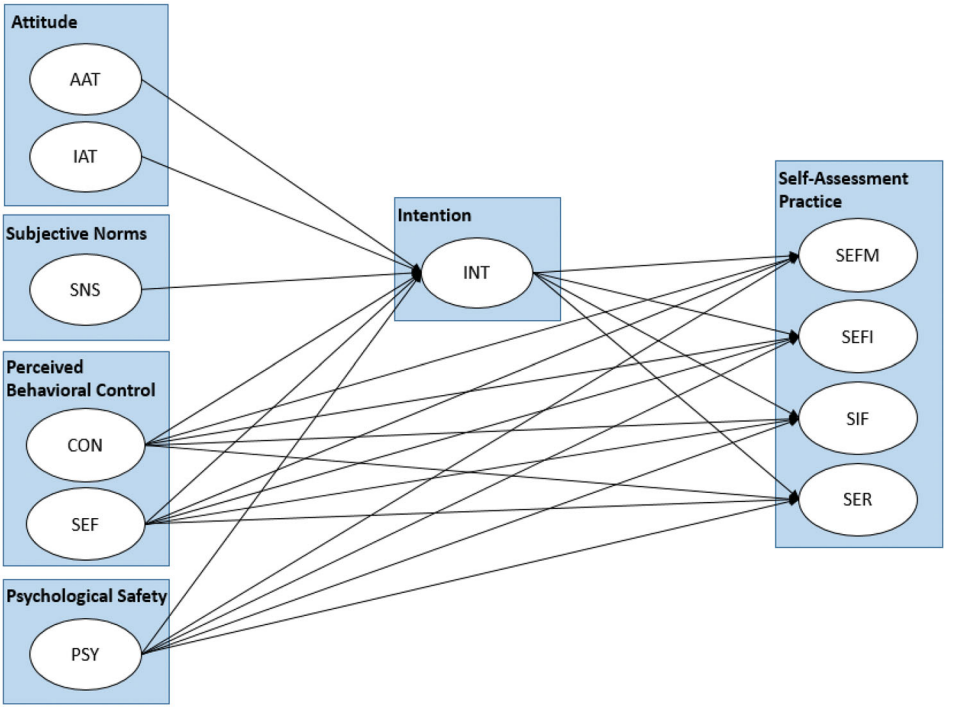


Figure 1. The extended model based on the Theory of Planned Behaviour.

this work) or to give weak reasons for their judgment (e.g. *I worked hard*) (Harris & Brown, 2013). Consequently, students' perceptions of psychological safety might be a crucial additional factor in determining students' self-assessment intentions and practices (Andrade & Brown, 2016). It may be that perceptions of psychological safety together with the subjective norms (i.e. perceived social pressure to perform or not perform the target behavior) determine how a self-assessment is formulated or expressed.

Therefore, the framework of TPB was extended by adding an additional component, psychological safety, in the present study (see Figure 1). From TPB, student attitudes toward self-assessment (i.e. affective, AAT; and instrumental, IAT) and their sense of subjective norms (SNS) were treated as predictors of intention (INT) to self-assess. Additionally, perceived behavioural control (i.e. controllability, CON; and self-efficacy, SEF) and psychological safety (PSY) predicted both intention to self-assess and self-assessment practices. This means intention acted as a partially mediated factor intervening between student attitudes and sense of norms and actual practices. Self-assessment practices were formulated as four correlated actions covering the whole process implied by self-assessment; that is, seeking external feedback through monitoring (SEFM) and inquiry (SEFI), from internal sources (SIF), and self-reflection (SER). To the extent that any of these self-assessment actions require interacting with

the teacher or classmates, it is legitimate to expect some potential impact of psychological safety.

The present study

The present study explored the predictors of students' self-assessment intentions and practices through a self-reported survey study. Figure 1, shows the hypothesised relationships:

H1: students' intentions to conduct self-assessment can be predicted by their attitude, subjective norms, perceived behavioural control, and psychological safety regarding self-assessment

H2: students' self-assessment practices can be predicted by their intentions, perceived behavioural control, and psychological safety regarding self-assessment.

Another important issue addressed in this paper was the need to validate recently developed measurement scales (i.e. Self-assessment Practice Scale, SaPS; Yan, 2018b) for self-assessment. The present study provides further evidence from an independent sample for the new scale.

Methods

Participants

A total of 1425 students (52% male) from 29 schools in Hong Kong were surveyed. Participants were normally aged between 9 and 15 years and enrolled in the final 3 years of primary schooling (Primary 4 to Primary 6) and the first 3 years of secondary schooling (Secondary 1 to Secondary 3) (Table 1). This study focused on students in the upper primary and lower secondary levels because (1) Primary 1–3 students may have difficulty in understanding the questionnaire items; and (2) secondary schools in Hong Kong are usually reluctant to release Secondary 4–6 students for research projects as they are facing the pressure of high-stakes examinations. It is worth noting that, in Hong Kong, self-assessment is a regular learning practice for Primary 4 to Secondary 3. For example, upon completing each learning unit, some schools require

Table 1. Demographics of participants.

	Frequency	Percentage (%)
School sector		
Primary	658	46.2
Secondary	767	53.8
Year level		
Primary 4	191	13.4
Primary 5	239	16.8
Primary 6	228	16.0
Secondary 1	248	17.4
Secondary 2	214	15.0
Secondary 3	305	21.4
Gender		
Male	737	51.7
Female	681	47.8
Missing	7	0.5
n	1425	100.0

students to fill in a self-assessment worksheet. Through the worksheet, students are guided to reflect their gains, areas that need further work, and the possible ways for improvement.

Instruments

TPB requires all constructs under investigation (e.g. attitude, subjective norms, etc.) to be compatible with the target behaviour (Ajzen, 2002), therefore it is a normal practice to use a specific set of scales for investigating a specific target behaviour. The scale development procedure and item crafting followed Ajzen's (2002) principles for TPB scale construction. The items were mostly modified from previous TPB scales on assessment (e.g. the Conceptions and Practices of Formative Assessment Questionnaire; Yan & Cheng, 2015) and the Conceptions of School-based Assessment Questionnaire; Yan, 2014). The new scales were subjected to intensive review by invited experts in the field of psychometrics and educational assessment and front-line teachers who had good understanding of students' self-assessment. The reviewers examined the item contents, ambiguities, anomalies, and possible bias in item wording. All items were administered in traditional Chinese script and students rated each item on a 6-point scale balanced ranging from 1 (Strongly Disagree) to 6 (Strongly Agree).

Self-assessment practices

The behaviour component was assessed with the SaPS (Yan, 2018b) which contains 20 items grouped into four subscales assessing four self-assessment actions including seeking external feedback through monitoring (SEFM) (5 items), seeking external feedback through inquiry (SEFI) (4 items), seeking internal feedback (SIF) (4 items) and self-reflection (SER) (7 items). These four actions correspond to two out of three components in Yan and Brown's (2017) self-assessment model including self-directed feedback seeking (SEFM, SEFI and SIF) and self-reflection (SER). The validation study found that the SaPS was a measure with satisfactory psychometric properties, having sufficiently robust internal estimates of reliability (i.e. Cronbach's α for the four subscales were 0.85, 0.84, 0.79, and 0.90, respectively). The EAP/PV reliabilities, obtained by dividing the variance of the individual expected a posteriori ability estimates by the estimated total variance of the latent ability (Wu, Adams, Wilson, & Haldane, 2007) were 0.88, 0.88, 0.80, and 0.90 respectively, indicating strong scale characteristics (Yan, 2018b).

Self-assessment TPB factors

For assessing the four traditional TPB components, six scales were developed. The AAT scale (7 items) gauged students' feelings or emotions aroused by self-assessment (e.g. Self-assessment is interesting). The IAT scale (12 items) examined students' appraisals of the consequences or purpose of self-assessment (e.g. Self-assessment helps me to understand my strengths and weaknesses). The SNS scale (7 items) tapped into the social norms by assessing students' perceptions of how important others regarded self-assessment (e.g. I believe my teachers want me to do self-assessment). The CON

scale (5 items) checked students' beliefs about their control when conducting self-assessment (e.g. I decide which method of self-assessment to use). The SEF scale (5 items) examined students' beliefs about their self-efficacy in conducting self-assessment (e.g. I have enough knowledge to implement self-assessment). The INT scale (6 items) investigates students' intention or willingness to conduct self-assessment (e.g. I willingly assess myself).

Psychological safety in self-assessment

The psychological safety scale (PSY) (8 items) was developed to assess whether students feel safe in both interpersonal (e.g. I feel uncomfortable if others know my self-assessment results) and personal aspects (e.g. It is a problem to accept my self-assessment results if they are below my expectations.) of self-assessment.

Data analysis

To establish the validity of the designed scales, each scale was analysed using Rasch analysis (Rasch, 1960). Rasch analysis converts the agreement ordinal rating scale onto a continuous interval scale which permits subsequent parametric analysis (e.g. *t*-test, ANOVA or regression analysis). Rasch analysis examines scale quality by checking the degree to which items in a scale reflect an underlying unidimensional latent construct; items with mean square (MNSQ) fit statistics falling in the range between 0.6 and 1.4 were retained (Smith, 2004). To construct the unidimensionality metrics, the Rasch model adopts a 'data fit the model' approach that requires the empirical data to satisfy a priori requirements essential for the purpose of fundamental measurement (Bond & Fox, 2015). Rather than relying on a point estimate of student location, Rasch modelling generates a range of plausible values to represent the reasonable range of student endorsement of latent abilities.

Since all constructs under investigation were theoretically related, a multidimensional Rasch model, instead of the conventional unidimensional Rasch model, was employed with these data. In a multidimensional Rasch model, each subscale measures a unidimensional latent construct and all subscales are calibrated simultaneously with their correlations taken into account. The Rasch rating scale model (RSM) (Andrich, 1978) was applied and the computer program *ConQuest* (Wu et al., 2007) was used for the analyses. When the psychometric properties of the scales were satisfactory, five sets of plausible values of the Rasch-calibrated person measures were obtained from *ConQuest*. This produced a manifest scale score for each person for each scale. This treatment permits path analysis using *Mplus* (Muthén & Muthén, 2017). While path analysis is normally associated with small sample studies, it is a legitimate way to approach manifest score values derived from Rasch modeling as opposed to using full latent trait modeling within structural equation modeling. Paths were created according to Figure 1 hypotheses.

As suggested (Fan & Sivo, 2007; Hooper, Coughlan, & Mullen, 2008; Hu & Bentler, 1998), the model-data fit was examined using the ratio of the χ^2 statistic to degrees of freedom, the comparative fit index (CFI), the gamma hat index, the root mean square error of approximation (RMSEA), and the standardised root mean square residual

(SRMR). Generally, models need not be rejected if the χ^2/df has $p > .05$, the CFI and gamma hat are $>.90$, and RMSEA and SRMR are $<.08$. Considering that gender and school level were reported having significant influence on students' self-assessment in previous studies (e.g. Panadero, Jonsson, & Botella, 2017; Yan, 2018a), path analyses with gender and school sector as the moderators on the relationships between intentions and the predictors were conducted in order to investigate whether the effect of attitude, subjective norms, perceived behavioural control and psychological safety on students' self-assessment practices via their intentions differ across gender and school sector (primary vs. secondary).

Results

Psychometric properties of the scales

Although the main purpose of this study is to explore factors that predict students' self-assessment intentions and practices, it is necessary to examine the scale quality at first as some of them were specifically developed for this study. Before analysis, all negatively-worded items were reverse coded so that a higher score stands for a higher level of latent traits (e.g. a more positive attitude, a higher level of self-efficacy, etc.). The results of Rasch analysis showed that a total of 9 items (3 items from AAT and SNS, 1 item each from IAT, CON and PSY) had statistically significant misfit. Not surprisingly, seven of the misfitting items were negatively-worded, which echoed previous findings that negatively-worded items cause problems (e.g. Brown, 2004; Dodeen, 2015; Molina, Rodrigo, Losilla, & Vives, 2014). Given that the scales still had sufficient items with adequate coverage of contents for subsequent analysis, all misfitting items were removed. Rasch analysis was re-conducted and the psychometric properties of the scales were satisfactory for use with the sample in the present study. Specifically, as shown in Table 2, the person separation reliabilities of the scales ranged between 0.76 and 0.93, indicating the items in each scale measured the target constructs well.

The final scales with item difficulties, outfit and infit statistics, and thresholds are reported in Table A1. The item difficulties ranged between −0.60 and 0.44 with very small standard error (between 0.03 and 0.04). All the retained items had satisfactory fit and well-ordered threshold values. Table A2 reports the Wright map of item and person location relationships. There is a good correspondence between student measures

Table 2. Number of items and reliabilities of the scales.

	Number of items		
Scale	Original scale	Final scale	Reliability
Affective attitude (AAT)	7	4	0.88
Instrumental attitude (IAT)	12	11	0.93
Subjective norms (SNS)	7	4	0.77
Controllability (CON)	5	4	0.82
Self-efficacy (SEF)	5	5	0.84
Psychological safety (PSY)	8	7	0.84
Intention (INT)	6	6	0.90
Seeking external feedback through monitoring (SEFM)	5	5	0.83
Seeking external feedback through inquiry (SEFI)	4	4	0.83
Seeking internal feedback (SIF)	4	4	0.76
Self-reflection (SER)	7	7	0.87

Table 3. Mean, standard deviation, and inter-correlations of person measures among scales.

Scale	1	2	3	4	5	6	7	8	9	10	11
1. AAT											
2. IAT	0.61*										
3. SNS	0.65*	0.63*									
4. CON	0.61*	0.59*	0.53*								
5. SEF	0.48*	0.50*	0.48*	0.51*							
6. PSY	0.37*	0.36*	0.30*	0.45*	0.28*						
7. INT	0.16*	0.18*	0.16*	0.19*	0.18*	0.23*					
8. SEFM	0.58*	0.48*	0.46*	0.49*	0.38*	0.25*	0.14*				
9. SEFI	0.52*	0.41*	0.43*	0.42*	0.35*	0.22*	0.16*	0.58*			
10. SIF	0.35*	0.27*	0.25*	0.30*	0.24*	0.22*	0.21*	0.32*	0.34*		
11. SER	0.63*	0.50*	0.46*	0.54*	0.42*	0.31*	0.14*	0.61*	0.55*	0.40*	
M	0.18	0.84	0.46	0.44	0.55	0.02	0.69	0.49	0.12	0.52	0.89
SD	2.42	2.07	1.40	1.85	1.65	1.26	2.30	1.51	1.67	1.47	1.70

Note. AAT: affective attitude scale; IAT: instrumental attitude scale; SNS: subjective norms scale; CON: controllability scale; SEF: self-efficacy scale; PSY: psychological safety scale; INT: intention scale; SEFM: seeking external feedback through monitoring; SEFI: seeking external feedback through inquiry; SIF: seeking internal feedback; SER: self-reflection. $^{*}p < .05$.

and item difficulties, suggesting that the items thresholds within each scale provides satisfactory measurement for the participating students.

Descriptive statistics

Table 3 provides the means and standard deviations of Rasch-calibrated person measures for each scale and the inter-correlations. Mean score values ranged from close to zero to as high as 0.89. It was found for the predictors, students had the least agreement with PSY scale (0.02) and the most agreement for IAT (0.84); while for self-assessment practices, students had least agreement for SEFI (0.12) and most agreement for SER (0.89). All the latent traits were positively and significantly correlated.

Primary analysis

The path model using the Rasch-calibrated measures as manifest variables had a satisfactory fit ($\chi^2 = 377.99$; $df = 16$; $\chi^2/df = 23.62$, $p = .000$; CFI = 0.91; gamma hat = 0.98; RMSEA = 0.082; SRMR = 0.072). The standardised regression coefficients are shown in Figure 2.

It was found that all predictors included in the model demonstrated significant impacts on the formation of intention. A further testing of the magnitude of the differences between the standardised parameter estimates of all predictors showed that the predicting power of instrumental attitude, affective attitudes, self-efficacy, and social norms on intention were significantly stronger than that of psychological safety and controllability. Together these predictors produced about 49% variance explained in the intention factor. The four self-assessment actions were strongly predicted by self-efficacy (mean regression weight = 0.36) and more moderately by intention (mean regression weight = 0.26). As indicated by the testing of the magnitude of the differences between the parameter estimates, the predicting power of self-efficacy was significantly stronger than that of intention for seeking external feedback through monitoring and self-reflection, but not for seeking external feedback through inquiry

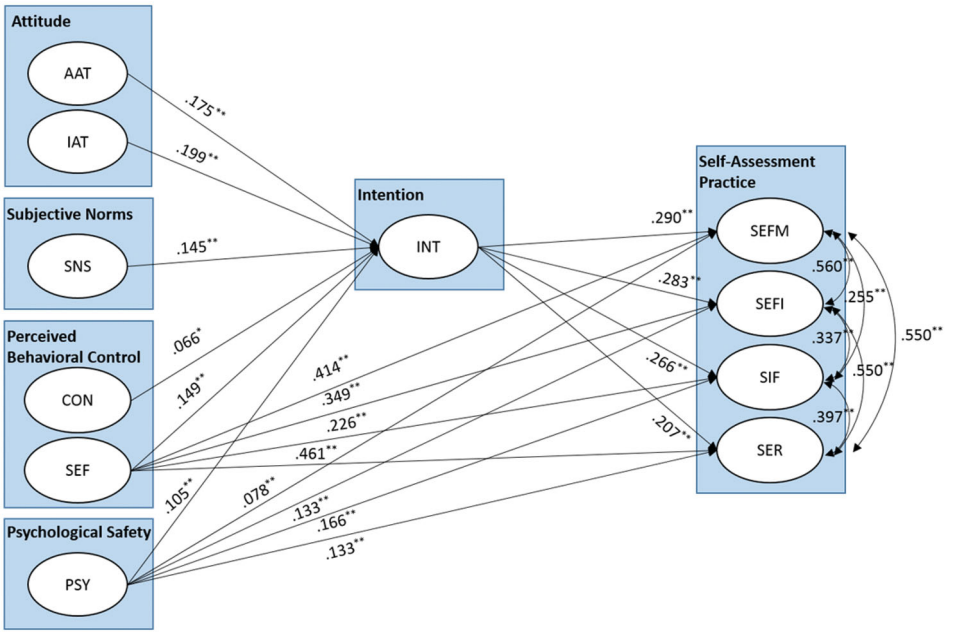


Figure 2. Path analysis based on the extended Theory of Planned Behaviour. Only significant paths are included in the figure. All predictors are correlated but not shown in this figure to avoid a messy presentation. AAT: affective attitude scale; IAT: instrumental attitude scale; SNS: subjective norms scale; CON: controllability scale; SEF: self-efficacy scale; PSY: psychological safety scale; INT: intention scale; SEFM: seeking external feedback through monitoring; SEFI: seeking external feedback through inquiry; SIF: seeking internal feedback; SER: self-reflection. $^{*}p < .05$; $^{**}p < .01$.

and seeking internal feedback. In contrast, controllability had no direct effect on self-assessment practices. The inter-correlations among the four self-assessment actions were weak to moderate ($r = 0.26$ to 0.56). The proportion of variance explained for the four self-assessment actions ranged from 29 to 46%.

The possible moderating role of gender and school sector was examined by path analyses with gender and school sector as the moderators on the relationships between intentions and the predictors. It was found that gender did not have significant moderating effect on the relationships between students' intentions and the predictors. In other words, the effects of attitude, subjective norms, perceived behavioral control and psychological safety on students' self-assessment practices via their intentions did not vary across gender. School sector was found to have a significant moderating effect on the path from subjective norms (-0.080, $p < .01$), self-efficacy (0.076, $p < .05$), and psychological safety (-0.043, $p < .05$) to students' intentions. Since the analyses set the parameter estimates for primary students as the baseline, these results mean that subjective norms and psychological safety had a stronger impact on intention for primary students, while self-efficacy had a stronger impact on intention for secondary students.

Discussion

The present study investigated the predictors of students' self-assessment intentions and practices under the framework of an extended theory of planned behaviour. The results generally supported the two proposed hypotheses; that is, students' intentions to conduct self-assessment can be predicted by their attitude, their subjective norms, their perceived behavioural control, and their sense of psychological safety regarding self-assessment (H1); and students' self-assessment practices can be predicted by their intentions, their perceived behavioural control, and their sense of psychological safety (H2).

Predictors of self-assessment intention

Among all the predictors of intention, instrumental attitude (IAT), affective attitude (AAT), self-efficacy (SEF), and subjective norms (SNS) were relatively strong predictors; while psychological safety (PSY) and controllability (CON) were relatively weak, but still significant, predictors. The crucial role of attitude in predicting behavioural intention echoes previous studies (Armitage & Conner, 2001; Cooke, Dahdah, Norman, & French, 2016). Instrumental attitude was found most powerful in predicting intention to conduct self-assessment although the differences were small. This replicates previous findings concerning the positive association of instrumental attitude with teacher intentions to administer formative assessment (Yan & Cheng, 2015) and school-based assessment (Yan, 2014). It appears that, for both students and teachers, the formation of the assessment behavioural intentions is largely influenced by the possible consequences associated with assessment practices. Nevertheless, the magnitudes of the prediction of all factors were relatively small (all regression coefficients were less than 0.2). This was in line with the results in Table 3 where all correlations with intention were small. It implies that the formation of self-assessment relies on the synergy of a variety of factors, rather than one or two determinant predictors.

It was found that gender had no significant moderating effect on the model's parameters. However, school sector was a moderator with significant influence on the predicting power of some variables. Subjective norms and psychological safety demonstrated a stronger impact in predicting intention to self-assess for primary students than for secondary students. In contrast, self-efficacy had a stronger impact on intention to self-assess for secondary students than for primary students. This finding has implications for instruction. Interventions attempting to promote student self-assessment could be developed from two different perspectives. One targets external factors, such as developing a supportive and safe learning environment that positively alters students' perceived norms and psychological safety within self-assessment. The other focuses on capacity building, such as enhancing students' self-efficacy for self-assessment. Although both perspectives may be essentially effective, the impact is likely to vary across different populations. The former might be more useful in primary classrooms while the latter may be more effective for secondary students.

Predictors of self-assessment practice

A unique feature of the present study is that it divided self-assessment practices into four actions (i.e. seeking external feedback through monitoring, seeking external feedback through inquiry, seeking internal feedback, and self-reflection), according to Yan and Brown's (2017) process model. Breaking down the self-assessment process into concrete actions and using them as the unit of analysis makes it possible to develop students' competence of these actions. Each step in the self-assessment process may have its own impact on the final self-assessment judgment and students may prioritise different aspects of the process (Yan, 2016). Therefore, it is meaningful to identify predictors of each self-assessment action and investigate whether there is differentiated predictive power of each predictor to different self-assessment action. Contrary to expectations, each one of these self-assessment actions was predicted more or less in the same way by self-efficacy, intentions, and psychological safety.

The greater influence of self-efficacy on self-assessment actions than intention does not correspond exactly to the TPB that assumes that intention has a stronger influence on the behaviour than perceptions of control (Ajzen, 1991). The direct measure of perceived behavioural control (i.e. controllability; CON) had a small, but statistically significant, impact on intention (INT) and had no significant impact on any of the self-assessment actions. This result corroborated the claim made in review studies (e.g. Cooke et al., 2016; Trafimow, Sheeran, Conner, & Finlay, 2002) and recent empirical studies (e.g. Brechan, 2016; Collins, Witkiewitz, & Larimer, 2011; Yan & Cheng, 2015) that controllability and self-efficacy were distinguishable constructs under the perceived behaviour control component of the TPB, and that self-efficacy had greater predictive power than controllability in predicting both intention and actual behaviour. Collins et al. (2011) have further argued that controllability might be less useful in predicting volitional behaviours (e.g. self-assessment) when internal determinants (e.g. confidence to implement self-assessment) influence a behaviour more than external factors (e.g. guidance to implement self-assessment). In this adapted model, self-efficacy was presumed to be legitimately an expression of perceived control, because self-efficacy has to do with confidence to perform a specific behaviour. The current results, however, show that confidence to perform self-assessment may be quite a different construct than perceptions of control. All students may have a sense of control, but having confidence in ability to perform the task may reflect a quite different psychological competence.

An important question thus arises with regard to in which way students' self-efficacy for self-assessment can be developed. Considerations on this issue may be necessary for instructional programmes that integrate self-assessment as an important curricular element. Previous studies demonstrated that self-efficacy can be developed and enhanced (Bandura, 1997; Jackson, 2002; van Dinther, Dochy, & Segers, 2011). For example, Bandura (1997) argued that self-efficacy arises from active mastery, vicarious experiences, verbal persuasion, and physiological/affective states. Accordingly, it is possible to enhance students' capacity and self-efficacy in making realistic self-assessments from different ways. For instance, teacher feedback on the validity of student self-assessments could be a verbal persuasion that is likely to enhance students' self-efficacy (Brown & Harris, 2014). With appropriate guidelines, having students

predict their performance (e.g. as an act of judgment of learning; Baars, Vink, Van Gog, de Bruin, & Paas, 2014) on a concrete task may also teach students through active mastery that their self-assessments are realistic. Teacher demonstrations of self-assessment actions on concrete and easy to follow tasks could provide vicarious experiences for building the capacity and confidence to realistically self-assess. The level of difficulty of tasks may need to be considered as students may gain confidence in resolving tasks of appropriate level of academic and emotional challenge. Demonstrations by somewhat more experienced peers may also be effective (Bandura, 1997). Empirical studies (Troop, Wallar, & Aspenlieder, 2015; Van Rooij, Jansen, & Van de Grift, 2017) have indicated that practice leads to increased self-efficacy because it increases familiarity with a particular behaviour. As Panadero et al. (2016) advocated in their discussion of a 'developmental approach to self-assessment', students need to have multiple opportunities to develop the skill of realistically evaluating their own work so that they can gain confidence in their capacity. Therefore, self-assessment activity should be regarded as routine in learning and instruction so that students have opportunity to engage in veridical or realistic self-assessment.

Another innovation of the present study is that it applied an extension of the TPB by adding psychological safety as a predictor of intention and behaviour regarding self-assessment. The results showed that psychological safety (PSY) had small but significant impact on all self-assessment actions. On one hand, psychological safety's impact on seeking external feedback means that it is an important *interpersonal variable*, as specified in peer assessment literature (e.g. Cheng & Tsai, 2012; Rotsaert, Panadero, Estrada, & Schellens, 2017). On the other hand, its impact on internal feedback seeking indicates that psychological safety is also an *intra-personal variable*. This is not surprising given that internally generated feedback is closely related to affective reactions (Butler & Winne, 1995). A safe and supportive learning environment may be necessary for self-assessment because such an environment encourages students to be honest about their own deficiencies or to take interpersonal risks, such as seeking feedback from other people (Brown & Harris, 2013). More importantly, a sense of internal psychological safety is crucial to protect students' self-esteem so that students can interpret unsatisfying self-assessment results as opportunities to learn rather than a proof of failure. Such internal psychological safety should be built on a firm shared classroom belief in the formative use of self-assessment in which self- assessment is viewed as a learning strategy rather than an assessment method (Harris & Brown, 2018). This belief will also encourage students to pursue realistic and verdict self-assessment so as to learn as much as possible.

In Hong Kong, however, the powerful role formal summative evaluations play suggests that an improvement-oriented approach to self-assessment may be subverted by the dominant view of assessment as a summative evaluation (Kennedy, Chan, & Fok, 2011). Such an approach might threaten the validity and integrity of self-assessment practices. Therefore, consideration of psychological safety is especially important in the context of Hong Kong. The inclusion of psychological safety within the TPB framework shows that Hong Kong school students are like students in other K-12 studies who were inclined to dissemble in self-assessments if they did not perceive psychological safety (Cowie, 2009; Harris & Brown, 2013; Raider-Roth, 2005; Ross, Rolheiser, & Hogaboam-Gray, 2002).

Implications for future research

While this study has provided original insights into the factors influencing students' self-assessment intentions and practices, further development in this line of research is necessary. The study relied on self-report data and, therefore, the results might be vulnerable to the mono-method bias and response bias (e.g. inaccurate memory and social desirability). For self-assessment practices, future studies could consider using direct measures (e.g. observations) to replace or substantiate self-reports. Although it is challenging to design objective or direct measures for attitude, intention, and inner processes regarding self-assessment, future research should employ innovative methods (e.g. digital traces, eye-tracking, etc.) to minimise the limitations associated with self-report data.

Given the important role self-efficacy plays in this model and the interaction of self-efficacy with actual demonstrated competence, it was unfortunate that a verifiable measure of actual competence was not available. Future studies would do well to obtain a reliable academic proficiency score as a control variable; it might be that the relations seen here will be different according to achievement. Furthermore, attention should be paid to the specificity of self-efficacy—it matters in which domain (e.g. mathematics vs. reading) one assesses one's ability. Hence, the academic proficiency measure and the domain in which the self-assessment occurs have to be aligned.

The Yan and Brown's model used in this study was developed in education but with teacher education students. This study seems to suggest that the statistical properties of that model hold with students in primary and secondary school. Nonetheless, it would be good to carry out exploratory studies with younger students to see how they formulate a self-assessment. It is likely they could follow a similar model because of the constraints of an outcome-based curriculum. It could be interesting to examine preschool students to ascertain how they formulate opinions about the quality of their learning or performances.

Differentiation of self-assessment practice into four actions brings in-depth insights into the self-assessment process, but its side effect is, to some extent, a violation of the requirement of specificity in the TPB (i.e. all components under investigation should refer to the same target behaviour; Ajzen, 2002). The self-assessment behaviour component was broken into four actions, but the other components referred to a general self-assessment practice. Based on the general picture provided by the present study, future research may narrow the scope and investigate each of the self-assessment actions individually within a TPB framework.

Another meaningful direction is to look for strategies to enhance the factors, identified in the present study, that promote self-assessment actions. Experimental studies with well-designed interventions will be particularly meaningful to find out whether and to what extent the particular self-assessment actions can be promoted. Furthermore, cultural factors should never be neglected in understanding self-assessment behaviour which is a human and internal psychological process. Cultural variations have been shown to have impact on factors relevant to self-assessment such as source of self-efficacy belief (Hendricks, 2016) and psychological safety in assessment (Brown & Harris, 2013). The model and hypotheses could be further tested in different cultural contexts.

Finally, the present study adopted a two-step analytical method which combined the Rasch model (with *ConQuest*) and path analysis (with *Mplus*). Although both Rasch analysis and structural equation modelling have been widely used in educational research, only a few studies have combined these methods in a meaningful context (e.g. Mok, Kennedy, & Zhu, 2013; Su & Wang, 2010; Yan & Cheng, 2015). In present study, the multidimensional Rasch analysis was applied instead of confirmatory factor analysis in conventional structural equation modelling due to the opportunities Rasch analysis provides in dealing with non-interval categorical raw scores and multiple subtests. In addition, utilising a set of plausible values instead of a single estimate of ability takes the measurement error of the ability estimate into account, leading to more accurate standard error of the coefficients in path analysis. Thus, taking advantage of both the Rasch model and path analysis or structural equation modeling warrants more discussion and application in the future.

Conclusions

In summary, the findings of the present study provide important insights for learning and instruction with regard to the factors that influence student self-assessment intentionality and behaviour. The extended TPB appears a useful framework for understanding self-assessment behaviour and self-efficacy plays the most important role in this framework. The present study lays a foundation for future investigations on how to promote meaningful self-assessment behaviour.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by a General Research Fund (GRF) from the Research Grants Council of Hong Kong [Project No.: EDUHK 18605715].

ORCID

Zi Yan http://orcid.org/0000-0001-9305-884X
Gavin T. L. Brown http://orcid.org/0000-0002-8352-2351

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, *50*(2), 179–211. doi:10.1016/0749-5978(91)90020-T

Ajzen, I. (2002). *Constructing a TpB questionnaire: Conceptual and methodological considerations*. Retrieved from https://pdfs.semanticscholar.org/0574/b20bd58130dd5a961f1a2db10fd1fcbae95d.pdf

Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education*, *32*(2), 159–181. doi:10.1080/02602930600801928

Andrade, H. L., & Brown, G. T. L. (2016). Student self-assessment in the classroom. In G. T. L. Brown and L. R. Harris (Eds.), *Handbook of human and social conditions in assessment* (pp. 319–334). New York, NY: Routledge.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, *43*(4), 561–573. doi:10.1007/BF02293814

Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *British Journal of Social Psychology*, *40*(4), 471–499. doi:10.1348/014466601164939

Baars, M., Vink, S., Van Gog, T., de Bruin, A., & Paas, F. (2014). Effects of training self-assessment and using assessment standards on retrospective and prospective monitoring of problem solving. *Learning and Instruction*, *33*, 92–107. doi:10.1016/j.learninstruc.2014.04.004

Baas, D., Castelijns, J., Vermeulen, M., Martens, R., & Segers, M. (2015). The relation between Assessment for Learning and elementary students' cognitive and metacognitive strategy use. *British Journal of Educational Psychology*, *85*(1), 33–46. doi:10.1111/bjep.12058

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). New York: Routledge.

Boud, D. (1999). Avoiding the traps: Seeking good practice in the use of self-assessment and reflection in professional courses. *Social Work Education*, *18*(2), 121–132. doi:10.1080/02615479911220131

Boud, D., Lawson, R., & Thompson, D. G. (2013). Does student engagement in self-assessment calibrate their judgement over time? *Assessment & Evaluation in Higher Education*, *38*(8), 941–956. doi:10.1080/02602938.2013.769198

Brechan, I. (2016). Travel intention: Relative value of transport alternatives. *Human Affairs*, *26*(4), 390–399. doi:10.1515/humaff-2016-0033

Brooks, V. (2002). *Assessment in secondary schools: The new teacher's guide to monitoring, assessment, recording, reporting and accountability*. Buckingham, UK: Open University Press.

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., Andrade, H., & Chen, F. (2015). Accuracy in student self-assessment: Directions and cautions for research. *Assessment in Education: Principles, Policy & Practice*, *22*(4), 444–457. doi:10.1080/0969594X.2014.996523

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 367–393). Thousand Oaks, CA: Sage.

Brown, G. T. L., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. *Frontline Learning Research*, *3*(1), 22–30. doi:10.14786/flr.v2i1.24

Burrus, J., & Moore, R. (2016). The incremental validity of beliefs and attitudes for predicting mathematics achievement. *Learning and Individual Differences*, *50*, 246–251. doi:10.1016/j.lindif.2016.08.019

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

Cassidy, S. (2007). Assessing 'inexperienced' students' ability to self-assess: Exploring links with learning style and academic personal control. *Assessment & Evaluation in Higher Education*, *32*(3), 313–330. doi:10.1080/02602930600896704

Cheng, K.-H., & Tsai, C.-C. (2012). Students' interpersonal perspectives on, conceptions of and approaches to learning in online peer assessment. *Australasian Journal of Educational Technology*, *28*(4), 599–618. doi:10.14742/ajet.830

Collins, S. E., Witkiewitz, K., & Larimer, M. E. (2011). The theory of planned behavior as a predictor of growth in risky college drinking. *Journal of Studies on Alcohol and Drugs*, *72*(2), 322–332. doi:10.15288/jsad.2011.72.322

Cooke, R., Dahdah, M., Norman, P., & French, D. P. (2016). How well does the theory of planned behaviour predict alcohol consumption? A systematic review and meta-analysis. *Health Psychology Review*, *10*(2), 148–167. doi:10.1080/17437199.2014.947547

Cowie, B. (2009). My teacher and my friends helped me learn: Student perceptions and experiences of classroom 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. 85–105). Charlotte, NC: Information Age Publishing.

Dodeen, H. (2015). The effects of positively and negatively worded items on the factor structure of the UCLA Loneliness Scale. *Journal of Psychoeducational Assessment*, *33*(3), 259–267. doi:10.1177/0734282914548325

Edmondson, A. C. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly*, *44*(2), 350–383. doi:10.2307/2666999

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, *42*(3), 509–529. doi:10.1080/00273170701382864

Fastré, G. M. J., van der Klink, M. R., Sluijsmans, D., & van Merriënboer, J. J. G. (2012). Drawing students' attention to relevant assessment criteria: Effects on self-assessment skills and performance. *Journal of Vocational Education & Training*, *64*(2), 185–198. doi:10.1080/13636820.2011.630537

Fife-Schaw, C., Sheeran, P., & Norman, P. (2007). Simulating behaviour change interventions based on the theory of planned behaviour: Impacts on intention and action. *British Journal of Social Psychology*, *46*(1), 43–68.

Hagger, M. S., Sultan, S., Hardcastle, S. J., & Chatzisarantis, N. L. D. (2015). Perceived autonomy support and autonomous motivation toward mathematics activities in educational and out-of-school contexts is related to mathematics homework behavior and attainment. *Contemporary Educational Psychology*, *41*, 111–123. doi:10.1016/j.cedpsych.2014.12.002

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

Harris, L. R., & Brown, G. T. L. (2018). *Using self-assessment to improve student learning*. New York, NY: Routledge.

Hendricks, K. S. (2016). The sources of self-efficacy: Educational research and implications for music. *Update: Applications of Research in Music Education*, *35*(1), 32–38. doi:10.1177/8755123315576535

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*(1), 53–60.

Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*(4), 424–453. doi:10.1037//1082-989X.3.4.424

Ibabe, I., & Jauregizar, J. (2010). Online self-assessment with feedback and metacognitive knowledge. *Higher Education*, *59*(2), 243–258. doi:10.1007/s10734-009-9245-6

Jackson, J. W. (2002). Enhancing self-efficacy and learning performance. *The Journal of Experimental Education*, *70*(3), 243–254. doi:10.1080/00220970209599508

Kennedy, K. J., Chan, J. K. S., & Fok, P. K. (2011). Holding policy-makers to account: Exploring 'soft' and 'hard' policy and the implications for curriculum reform. *London Review of Education*, *9*(1), 41–54. doi:10.1080/14748460.2011.550433

Kissling, E. M., & O'Donnell, M. E. (2015). Increasing language awareness and self-efficacy of FL students using self-assessment and the ACTFL proficiency guidelines. *Language Awareness*, *24*(4), 283–302. doi:10.1080/09658416.2015.1099659

McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons*, *87*(1), 40–49.

Mok, M. M. C., Kennedy, K. J., & Zhu, J. (2013). Adolescents' civic engagement: A study of five Asian societies. In G. A. D. Liem, & A. Bernardo (Eds.), *Advancing cross-cultural perspectives on educational psychology: A Festschrift for Dennis McInerney* (pp. 168–192). Charlotte, NC: Information Age Publishing.

Mok, V., & Lee, A. Y. K. (2013). A case study on application of the Theory of Planned Behaviour: Predicting physical activity of adolescents in Hong Kong. *Journal of Community Medicine & Health Education*, *3*(5), 231–236. doi:10.4172/2161-0711.1000231

Molina, J. G., Rodrigo, M. F., Losilla, J.-M., & Vives, J. (2014). Wording effects and the factor structure of the 12-item General Health Questionnaire (GHQ-12). *Psychological Assessment*, *26*(3), 1031–1037. doi:10.1037/a0036472

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide (Version 8.0) [Computer software and manual]*. Los Angeles, CA: Muthén & Muthén.

Panadero, E., & Alonso-Tapia, J. (2013). Self-assessment: Theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students. *Electronic Journal of Research in Educational Psychology*, *11*(2), 551–576.

Panadero, E., Alonso-Tapia, J., & Antonio, H. J. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, *22*(6), 806–813. doi:10.1016/j.lindif.2012.04.007

Panadero, E., Brown, G. T., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, *28*(4), 803–830. doi:10.1007/s10648-015-9350-2

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, *22*, 74–98.

Perugini, M., & Bagozzi, R. P. (2001). The role of desires and anticipated emotions in goal-directed behaviors: Broadening and deepening the theory of planned behavior. *British Journal of Social Psychology*, *40*(1), 79–98. doi:10.1348/014466601164704

Raider-Roth, M. B. (2005). Trusting what you know: Negotiating the relational context of classroom life. *Teachers College Record*, *107*(4), 587–628.

Rasch, G. (1960). *Probabilistic models for some intelligence and achievement test*. Copenhagan, Denmark: Danish Institute for Educational Research.

Rivis, A., Sheeran, P., & Armitage, C. J. (2009). Expanding the affective and normative components of the theory of planned behavior: A meta-analysis of anticipated affect and moral norms. *Journal of Applied Social Psychology*, *39*(12), 2985–3019. doi:10.1111/j.1559-1816.2009.00558.x

Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (2002). Influences on student cognitions about evaluation. *Assessment in Education: Principles, Policy & Practice*, *9*(1), 81–95. doi:10.1080/09695940220119201

Rotsaert, T., Panadero, E., Estrada, E., & Schellens, T. (2017). How do students perceive the educational value of peer assessment in relation to its social nature? A survey study in Flanders. *Studies in Educational Evaluation*, *53*, 29–40. doi:10.1016/j.stueduc.2017.02.003

Sargeant, J., Armson, H., Chesluk, B., Dornan, T., Eva, K., Holmboe, E., … van der Vleuten, C. (2010). The processes and dimensions of informed self-assessment: A conceptual model. *Academic Medicine*, *85*(7), 1212–1220. doi:10.1097/ACM.0b013e3181d85a4e

Siegesmund, A. (2016). Increasing student metacognition and learning through classroom-based learning communities and self-assessment. *Journal of Microbiology & Biology Education*, *17*, 204–214. doi:10.1128/jmbe.v17i2.954

Smith, R. M. (2004). Fit analysis in latent trait measurement models. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch measurement: Theory, models and applications* (pp. 73–92). Maple Grove, MN: JAM Press.

Stanec, A. D. S. (2009). The Theory of Planned Behavior: Predicting teachers' intentions and behavior during fitness testing. *Journal of Teaching in Physical Education*, *28*(3), 255–271. doi:10.1123/jtpe.28.3.255

Su, C.-M., & Wang, W.-C. (2010). Nonlinear structural equation models: An item response modeling approach to categorical response variables. *Journal of Educational Research and Development*, *6*, 1–45.

Tai, J., Ajjawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: Enabling students to make decisions about the quality of work. *Higher Education*, *76*(3), 467–481. doi:10.1007/s10734-017-0220-3

Trafimow, D., Sheeran, P., Conner, M., & Finlay, K. A. (2002). Evidence that perceived behavioral control is a multidimensional construct: Perceived control and perceived difficulty. *British Journal of Social Psychology*, *41*(1), 101–121. doi:10.1348/014466602165081

Troop, M., Wallar, L., & Aspenlieder, E. (2015). Developing graduate students' self-efficacy with learner-centred lecturing. *Canadian Journal of Higher Education*, *45*(3), 15–33.

Tucker, A. L., Nembhard, I. M., & Edmondson, A. (2007). Implementing new practices: An empirical study of organizational learning in hospital intensive care units. *Management Science*, *53*(6), 894–907. doi:10.1287/mnsc.1060.0692

van Dinther, M., Dochy, F., & Segers, M. (2011). Factors affecting students' self-efficacy in higher education. *Educational Research Review*, *6*(2), 95–108. doi:10.1016/j.edurev.2010.10.003

Van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2009). Peer assessment for learning from a social perspective: The influence of interpersonal variables and structural features. *Educational Research Review*, *4*(1), 41–54.

Van Gennip, N. A. E., Segers, M. S. R., & Tillema, H. H. (2010). Peer assessment as a collaborative learning activity: The role of interpersonal factors and conceptions. *Learning and Instruction*, *20*(4), 280–290. doi:10.1016/j.learninstruc.2009.08.010

Van Rooij, E. C. M., Jansen, E. P. W. A., & Van de Grift, W. J. C. M. (2017). Factors that contribute to secondary school students' self-efficacy in being a successful university student. *Research in Post-Compulsory Education*, *22*(4), 535–555. doi:10.1080/13596748.2017.1381301

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, A. A. (2007). *ACER ConQuest [Computer program and manual]*. Victoria, Australia: ACER Press.

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. doi:10.1080/13803611.2013.877394

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. (2018a). Student self-assessment practices: The role of gender, year level, and goal orientation. *Assessment in Education: Principles, Policy & Practice*, *25*(2), 183–199. doi:10.1080/0969594X.2016.1218324

Yan, Z. (2018b). The Self-assessment Practice Scale (SaPS) for students: Development and psychometric studies. *The Asia-Pacific Education Researcher*, *27*(2), 123–135. doi:10.1007/s40299-018-0371-8

Yan, Z. (2019). Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assessment & Evaluation in Higher Education*. Advance online publication. doi:10.1080/02602938.2019.1629390

Yan, Z., & Brown, G. T. L. (2017). A cyclical self-assessment process: Towards a model of how students engage in self-assessment. *Assessment & Evaluation in Higher Education*, *42*(8), 1247–1262. doi:10.1080/02602938.2016.1260091

Yan, Z., & Cheng, E. C. K. (2015). Primary teachers' attitudes, intentions and practices regarding formative assessment. *Teaching and Teacher Education*, *45*, 128–136.

Yan, Z., & Sin, K. F. (2014). Inclusive education: Teachers' intentions and behaviour analyzed from the viewpoint of the Theory of Planned Behaviour. *International Journal of Inclusive Education*, *18*(1), 72–85. doi:10.1080/13603116.2012.757811

Yan, Z., & Sin, K. F. (2015). Exploring the intentions and practices of principals regarding inclusive education: An application of the Theory of Planned Behaviour. *Cambridge Journal of Education*, *45*(2), 205–221. doi:10.1080/0305764X.2014.934203

Zhan, Y., Fang, Y., Wei, K.-K., & Chen, H. (2010). Exploring the role of psychological safety in promoting the intention to continue sharing knowledge in virtual communities. *International Journal of Information Management*, *30*(5), 425–436. doi:10.1016/j.ijinfomgt.2010.02.003

Appendix

Table A1. Difficulties, outfit and infit statistics, and thresholds of the items in the final scales.

Scale	Item	Difficulty	S.E.	Outfit MNSQ	Infit MNSQ	τ_1	τ_2	τ_3	τ_4	τ_5
AAT	1. I like self-assessment.	0.07	0.04	1.00	0.99	−2.60	−1.95	−0.46	1.67	3.34
	2. Self-assessment is an engaging experience for me.	−0.26	0.04	0.91	0.90					
	3. Self-assessment is interesting.	−0.02	0.04	0.96	0.92					
	4. I enjoy the process of self-assessment.	0.21	0.04	1.08	1.09					
IAT	1. Self-assessment gives an accurate appraisal of my performance.	−0.12	0.04	1.13	1.10	−2.27	−1.88	−0.64	1.31	3.48
	2. Self-assessment helps me to understand my strengths and weaknesses.	−0.52	0.04	1.27	1.26					
	3. Self-assessment raises my interest in learning.	0.29	0.04	0.96	0.94					
	4. Self-assessment encourages me to work harder.	0.16	0.04	0.81	0.82					
	5. Self-assessment encourages me to be independent in my learning.	0.17	0.04	0.93	0.91					
	6. Self-assessment improves my confidence in learning.	0.13	0.04	0.93	0.93					
	7. Self-assessment helps me learn more efficiently.	0.16	0.04	0.82	0.81					
	8. Self-assessment raises my scores and grades.	0.25	0.04	0.95	0.93					
	9. Self-assessment helps me track my progress.	−0.06	0.04	1.00	1.01					
	10. Self-assessment is a way to determine how much I have learned from teaching.	−0.24	0.04	1.02	1.04					
	11. Self-assessment helps me check my progress against achievement objectives.	−0.22	0.04	1.10	1.11					
SNS	1. I believe the principal of my school wants all students to self-assess.	−0.25	0.03	1.06	1.08	−1.27	−1.39	−0.47	1.11	2.03
	2. I believe my teachers want me to do self-assessment.	−0.24	0.03	0.81	0.82					
	3. Among my friends we know self-assessment will help us learn.	0.44	0.03	0.90	0.89					
	4. I believe my classmates take self-assessment seriously.	0.05	0.03	1.16	1.18					
CON	1. I have the freedom whether or not to implement self-assessment.	−0.20	0.03	1.07	1.08	−1.65	−1.52	−0.74	1.20	2.71
	2. The frequency of self-assessment is up to me.	0.11	0.03	0.82	0.80					
	3. I decide which method of self-assessment to use.	−0.07	0.03	0.78	0.77					
	4. I am in charge of deciding when I should do self-assessment.	0.16	0.03	1.29	1.28					
SEF	1. I know how to implement self-assessment.	0.12	0.03	0.98	0.97	−1.72	−1.62	−0.61	1.22	2.74

(*continued*)

Table A1. Continued.

Scale	Item	Difficulty	S.E.	Outfit MNSQ	Infit MNSQ	τ_1	τ_2	τ_3	τ_4	τ_5
	2. I can find materials (e.g. reference books, exam papers) against which to assess myself.	0.22	0.03	1.07	1.05					
	3. I have enough knowledge to implement self-assessment.	−0.01	0.03	0.82	0.81					
	4. I can design appropriate tasks to assess myself.	0.21	0.03	0.82	0.81					
	5. I can tell when my work is actually good or bad.	−0.53	0.03	1.26	1.30					
PSY	1. I feel uncomfortable if others know my self-assessment results.	0.37	0.03	1.32	1.34	−1.49	−0.91	0.27	0.99	1.14
	2. It is difficult to accept my self-assessment results if they are below my expectations.	0.09	0.03	0.92	0.90					
	3. I blame myself if my self-assessment is not positive.	0.05	0.03	1.00	1.00					
	4. I am too embarrassed to ask others for feedback in the process of self-assessment.	0.00	0.03	0.83	0.83					
	5. Honest self-assessment brings me negative consequences.	−0.37	0.03	1.05	1.04					
	6. Others will think that I overvalue myself if I give myself high marks in my self-assessment.	0.06	0.03	1.04	1.01					
	7. Others will think badly of me if my self-assessment results are poor.	−0.19	0.03	0.93	0.92					
INT	1. I willingly assess myself.	−0.21	0.04	1.33	1.38	−2.32	−1.78	−0.78	1.55	3.32
	2. Of my own accord, self-assessment is integrated into my learning.	−0.01	0.04	0.83	0.83					
	3. I am enthusiastic about making sure self-assessment is part of my studying.	0.09	0.04	0.87	0.87					
	4. I am willing to design appropriate assessment tasks for self-assessment.	0.11	0.04	0.90	0.90					
	5. I readily make an effort to assess myself.	−0.13	0.04	0.92	0.94					
	6. I plan to implement self-assessment in all my studies.	0.16	0.04	1.03	1.02					
SEFM	1. I check whether I have mastered the course content by doing extra exercises.	−0.17	0.03	0.86	0.88	−1.73	−1.22	−0.62	0.90	2.66
	2. I check whether I have fully understood the course content by doing past exam papers.	−0.27	0.03	0.81	0.82					
	3. I keep track of my progress by recording my performance.	0.29	0.03	1.03	1.01					
	4. I ask myself questions in my head to check whether I have understood the course content.	−0.08	0.03	0.88	0.90					

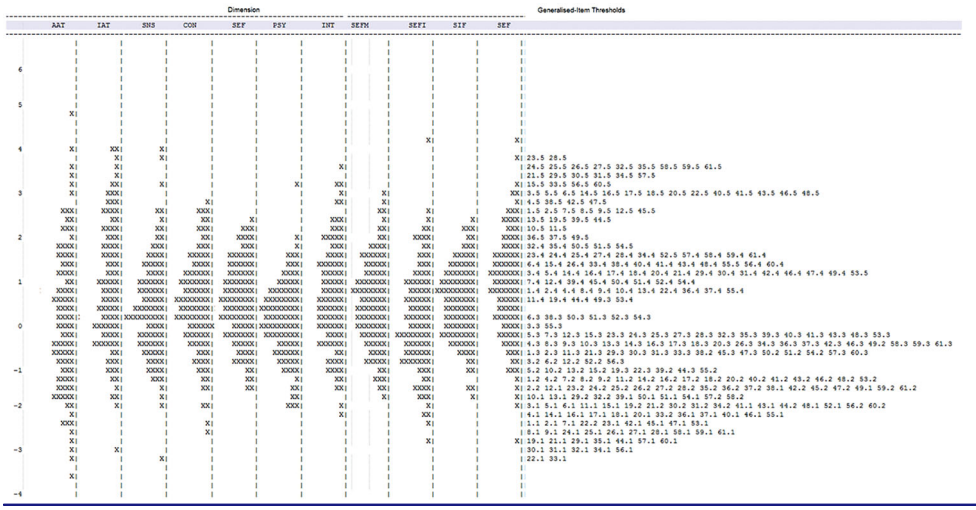
(*continued*)

Table A1. Continued.

Scale	Item	Difficulty	S.E.	Outfit MNSQ	Infit MNSQ	τ_1	τ_2	τ_3	τ_4	τ_5
	5. I check my performance against the answers in text books or on websites.	0.23	0.03	1.34	1.33					
SEFI	1. I ask my teachers to give me feedback about my performance.	0.35	0.03	0.93	0.92	−1.97	−1.17	−0.33	0.91	2.56
	2. I ask my family members to give me advice on my work.	−0.02	0.03	1.10	1.15					
	3. I ask my friends to tell me how to improve my learning.	−0.16	0.03	0.95	0.97					
	4. I ask my fellow group members to evaluate my contributions to group work tasks.	−0.16	0.03	0.92	0.94					
SIF	1. My gut feelings tell me whether my work is good or bad.	−0.10	0.03	1.03	1.05	−1.10	−1.19	−0.70	0.79	2.20
	2. My emotions influence my evaluation on my learning performance.	−0.20	0.03	1.31	1.36					
	3. How my body feels tells me how well I am doing.	0.25	0.03	0.75	0.75					
	4. My intuition tells me if I am doing a good job or not.	0.05	0.03	0.83	0.84					
SER	1. I seek out the reasons for mistakes I made after getting back marked work.	0.03	0.03	1.01	1.00	−1.77	−1.39	−0.72	1.08	2.80
	2. I think about how much sense the comments of other people (e.g. teachers, family members, and friends) regarding my work make to me.	0.26	0.03	1.25	1.23					
	3. Any areas I am unsure of after finishing my work, I go over again.	0.03	0.03	0.97	0.99					
	4. As I study, I think about whether the way I am studying is really helping me learn.	0.12	0.03	0.92	0.89					
	5. When I do exercise, I look at what I got wrong or did poorly on to guide me as to what I should learn next.	0.09	0.03	0.82	0.80					
	6. I pay attention to my assessment results in order to identify what I can do better next time.	−0.60	0.04	1.00	1.04					
	7. I reflect on my weaknesses when I discuss study-related issues with my classmates.	0.07	0.03	1.00	1.00					

Note. AAT: affective attitude scale; IAT: instrumental attitude scale; SNS: subjective norms scale; CON: controllability scale; SEF: self-efficacy scale; PSY: psychological safety scale; INT: intention scale; SFM: feedback through monitoring; SFI: feedback through inquiry; SIF: internal feedback; SER: self-reflection; S.E.: standard error; τ: threshold parameters.

Table A2. Wright-map of the scales.



Note. AAT: affective attitude scale; IAT: instrumental attitude scale; SNS: subjective norms scale; CON: controllability scale; SEF: self-efficacy scale; PSY: psychological safety scale; INT: intention scale; SFM: feedback through monitoring; SFI: feedback through inquiry; SIF: internal feedback; SER: self-reflection.