


Validation of a Subject-Specific Student Self-Assessment Practice Scale (SaPS) Among Secondary School Students in the Philippines

Journal of Psychoeducational Assessment
2021, Vol. 0(0) 1–13
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0734282921994374
journals.sagepub.com/home/jpa


Norman B. Mendoza¹ , and Zi Yan¹ 

Abstract

Self-assessment is fundamental to self-regulated learning; however, instruments to measure self-assessment practices are limited to a few developed educational systems. This study examined the psychometric properties of the Self-assessment Practice Scale (SaPS) in the English language subject using data from 778 secondary school students from the Philippines. We used confirmatory factor analysis (CFA) and Rasch analysis to test the SaPS' within-network validity, then bivariate correlations and structural equation modeling (SEM) for between-network validity. The CFA supported the scale's four-factor structure, and the Rasch analysis supported the scale's dimensionality, rating scale effectiveness, and item fit statistics. The four SaPS subscales were positively correlated to agentic, cognitive, and metacognitive engagement. SEM results show that all SaPS factors (except self-monitoring) had significant associations to the engagement outcomes. This study highlights the sound psychometric properties of SaPS in a new educational context and its applicability as a subject-specific measure of assessment-as-learning strategies.

Keywords

Self-assessment Practice Scale, validation, Philippines, self-regulated learning, English language subject

Self-assessment is generally conceptualized as a process through which students judge their performance based on the information and evidence collected from different sources (Boud, 1995; McMillan & Hearn, 2008; Yan & Brown, 2017). Broadly speaking, self-assessment occurs when “students reflect on the quality of their work, judge the degree to which it reflects explicitly stated goals or criteria, and revise accordingly” (Andrade & Valtcheva, 2009, p. 13). Self-assessment is a core skill for self-regulated learning (Panadero & Alonso-Tapia, 2013; Yan, Brubacher, Boud, & Powell, 2020) and lifelong learning (Papanthymou, 2018; Siegesmund, 2017; Yan & Brown, 2017). It is linked with key learning outcomes, such as school achievement (McDonald & Boud, 2003; Yan, Chiu, & Ko, 2020), student motivation (McMillan & Hearn, 2008), and self-regulated learning (Panadero, Jonsson, & Botella, 2017), among others (see Andrade, 2019; Brown & Harris, 2013,

¹The Education University of Hong Kong, Hong Kong

Corresponding Author:

Norman B. Mendoza, Department of Curriculum and Instruction, Faculty of Education and Human Development, The Education University of Hong Kong, D1-IF, Lo Ping Road, Tai Po, Hong Kong SAR, Hong Kong HKG, China.
Email: normanmendoza0421@gmail.com

for reviews). Because of the positive association between self-assessment and key learning outcomes, researchers have recognized the need to understand the self-assessment process and develop instruments to assess self-assessment (Panadero, Brown, & Strijbos, 2016; Yan & Brown, 2017). Doing so can help in programs or interventions designed to enact and maximize the impact of self-assessment.

Yan (2018a) adopted a theory-driven approach in developing the Self-assessment Practice Scale (SaPS) in the Hong Kong school context, which has been successfully applied in other contexts, such as higher education (Yan, 2020b) and professional training in western culture (Yan, Brubacher, Boud, & Powell, 2020). The instrument evaluates students' use of two core self-assessment processes: seeking self-directed feedback and self-reflection. However, the SaPS' validity for evaluating learner-centered formative assessment has not yet been investigated in developing countries with varied educational curricula (e.g., The Philippines). Validating the SaPS in the Philippines is especially crucial and beneficial, given the Philippines' poor performance in its first participation in the recent Program for International Student Assessment or PISA 2018 (OECD, 2019).

This study aims to evaluate the psychometric properties of the SaPS (Yan, 2018a) in the Philippines. This research is a first attempt to investigate the validity of the SaPS in the English language subject at a secondary school level. This could then extend the use of the SaPS across cultures in light of improving the use of 21st-century skills like student self-assessment.

SaPS

Assessing self-assessment *practices* is different from evaluating self-assessment *accuracy*. Although there is a consensus that self-assessment is a complex process incorporating multiple steps (see Andrade Du & Wang, 2008; Boud, 1995), the understanding of self-assessment practices (i.e. the actions students do during the self-assessment process) is surprisingly limited. Yan and Brown (2017) proposed a cyclical process model of self-assessment that explicitly outlines three sequential actions of self-assessment including (1) determining assessment criteria, (2) self-directed feedback seeking, and (3) engaging in self-reflection. These actions were included in self-report measures following the cyclical process of self-assessment (Yan, 2016, 2018b), which evolve into the SaPS (Yan, 2018a). To our knowledge, the SaPS is the only available theory-driven instrument that is specifically designed for assessing self-assessment practices.

The SaPS is focused on the self-assessment processes of self-directed feedback seeking and self-reflection. The scale starts with "*When I study...*," followed by self-assessment items which are grouped into four subscales corresponding to the four self-assessment actions: seeking external feedback through monitoring (SEFM; 5 items), seeking external feedback through inquiry (SEFI; 4 items), seeking internal feedback (SIF; 4 items), and engaging in self-reflection (SR; 7 items). The scale was validated with a sample of 2906 primary and secondary Hong Kong students. The results of factor analysis and Rasch analysis for SaPS supported its psychometric properties and structural validity (Yan, 2018a).

The SaPS and its short form and adapted versions have been used for graduate students (Yan, 2020b) and professional trainees in a Western cultural context (Yan, Brubacher, Boud, & Powell, 2020). The scale can track the development of self-assessment skills and can be used to inform teaching strategies that can promote self-assessment for optimizing learning outcomes. Evidence from recent interventions (e.g., randomized control trials and self-diaries) points to the effectiveness of self-assessment practices in increasing student achievement, self-regulation, and motivation (Meusen-Beekman, Joosten-ten Brinke, & Boshuizen, 2016; Yan, Chiu, & Ko, 2020).

The Current Study

This study extends the use of the SaPS as a critical tool in evaluating students' self-assessment practices in the English language as a specific subject domain. We adopted the network construct validation approach (see [Martin, 2007](#); [Martin & Marsh, 2006](#)) which involves examining the scale's within-network construct validity (i.e. factor structure and internal reliability) and between-network construct validity (i.e. the association of SaPS' dimensions with criterion-related constructs).

We used agentic, cognitive, and metacognitive engagement outcomes as correlates for testing the SaPS' between-network validity since these outcomes are viewed as self-regulatory learning (SRL) strategies ([Fredricks, Blumenfeld, & Paris, 2004](#); [Zimmerman & Schunk, 2004](#)) similar to self-assessment (see [Brown & Harris, 2013](#); [Panadero et al., 2017](#)). Agentic engagement is defined as "students' constructive contribution into the flow of the instruction they receive [in school]" (e.g., asking questions, communicating their thoughts and needs; [Reeve & Tseng, 2011](#), p. 258) and is conceptualized as a key component of overall student engagement ([Reeve, Cheon, & Jang, 2020](#)). Cognitive engagement is composed of internal indicators of SRL, such as students' striving and effort to understand complex ideas and master difficult skills ([Fredricks et al., 2004](#); [Fredricks & McColskey, 2012](#)). Metacognitive engagement consists of SRL strategies such as planning, monitoring, and revising one's schoolwork ([Wolters, 2004](#)). Metacognitive engagement has been posed as integrating motivation and SRL (see [Zimmerman & Moylan, 2009](#)) and is associated with self-assessment ([Siegesmund, 2017](#)). These student engagement outcomes are therefore posited as constructs interrelated with self-assessment.

Further, we aim to extend the utility of SaPS as a subject-specific instrument since students might employ different learning strategies (e.g., self-assessment practices and engagement outcomes) in different subject domains in school (e.g., [Wigfield, Eccles, Mac Iver, Reuman, & Midgley, 1991](#); [Wigfield, Guthrie, Tonks, & Perencevich, 2004](#)). Therefore, we modified the opening SaPS prompt from "When I study" to "When I learn English..." to frame subject specificity. Such a procedure has been used previously with closely related constructs (e.g., self-concept and self-efficacy; [Lent, Brown & Gore, 1997](#)). Contextualizing self-assessment practices has also been recommended in a separate validation study ([Yan, Brubacher, Boud, & Powell, 2020](#)). The engagement outcomes measured are also framed in the context of the English language subject.

Method

Participants

Participants in the study were 778 secondary school students from the Philippines¹. Participants with missing data greater than 5% and outliers for each of the measures were removed (see Data analysis section for details). Data from 673 students remained for the final analysis. The data consist of 186, 158, 157, and 172 students from Grades 7, 8, 9, and 10, respectively. Although the students' age ranged from 11 to 19 years old ($M = 14.14$, $SD = 1.51$), there was only one 11-year-old and 12 18- and 19-year-olds. More than half of the participants were females ($n = 376$, 55.87%). It was vital to have a balanced number of participants from each grade level since SRL strategies can be developmentally influenced (see [Paris & Newman, 1990](#)).

Procedures

Data were collected through a paper-and-pen survey method. Procedures for this study were approved by the Human Research Ethics Committee of both authors' affiliated university. A

Table 1. Bivariate correlations and descriptive statistics (n = 673).

	SaPS				Engagement		
	1	2	3	4	5	6	7
1. SEFM (monitoring)	(0.72)						
2. SEFI (inquiry)	0.54 ^{***}	(0.76)					
3. SIF (internal feedback)	0.53 ^{***}	0.44 ^{***}	(0.68)				
4. SR (reflection)	0.66 ^{***}	0.56 ^{***}	0.59 ^{***}	(0.81)			
5. Agentic	0.55 ^{***}	0.57 ^{***}	0.42 ^{***}	0.59 ^{***}	(0.87)		
6. Cognitive	0.56 ^{***}	0.44 ^{***}	0.48 ^{***}	0.68 ^{***}	0.67 ^{***}	(0.84)	
7. Metacognitive	0.55 ^{***}	0.42 ^{***}	0.54 ^{***}	0.64 ^{***}	0.62 ^{***}	0.71 ^{***}	(0.73)

Note. SEFM: seeking external feedback through monitoring; SEFI: seeking external feedback through inquiry; SIF: seeking internal feedback. * $p < .05$. ** $p < .01$. *** $p < .001$. Values shown in parentheses on diagonal are internal consistency reliabilities of the scales (Cronbach's alpha).

research assistant from the Philippines administered the questionnaires during the students' class hours and was present during the data collection to respond to potential student inquiries. The questionnaires were in the English language as English is the medium of instruction in the Philippines' K-12 schools. The students took about 10 minutes to complete the questionnaire.

Measures

SaPS. The SaPS (Yan, 2018a) is a 20-item instrument based on the cyclical model of the self-assessment process (Yan & Brown, 2017). It comprises four subscales: seeking external feedback through monitoring, seeking external feedback through inquiry, seeking internal feedback and self-reflection (see Table 1 for scale items). The internal consistency of the subscales ranged from $\alpha = 0.69$ to 0.81 (see Table 2).

Engagement outcomes. We used a 5-item *Agentic Engagement Scale* (Reeve & Tseng, 2011) which consists of items on students' constructive contribution to the flow of the instruction and feedback they receive in school (e.g., "During English class, I ask questions to help me learn."). We used Wolters's (2004) learning strategies questionnaire, which was derived from the Motivated Strategies for Learning Questionnaire (Pintrich, Smith, Garcia, & Mckeachie, 1993), to assess *cognitive engagement* (4 items) and *metacognitive engagement strategies* (4 items). Cognitive engagement items consist of elaboration-based learning strategies (e.g., "I adjust whatever we are learning so I can learn as much as possible."), while metacognitive engagement items include self-regulation strategies (e.g., "Before I begin to study, I think about what I want to get done."). These instruments' internal consistencies in the present study are $\alpha = 0.87$, 0.84, and 0.73, respectively.

Data Analysis

For the within-network validity, confirmatory factor analysis (CFA) and Rasch analysis were used. Numerous empirical studies have applied this complementary combined analytical approach (e.g., Chang & Engelhard, 2016; Testa et al., 2019; West et al., 2020; Yan, 2018a; Yan, Brubacher, Boud, Powell et al., 2020) to examine unique psychometric information of the scale items and subscales. The structural validity of the four-factor model of SaPS was assessed using CFA, where $n = 19$ participants with more than 5% of item-level missing data were excluded (Tabachnick, Fidell, & Ullman, 2007). Responses with item-level missing data were imputed using multiple

Table 2. Item difficulties, standard errors, and item fit statistics for the SaPS items ($n=768$).

Scale/Item	Item Measure*	SE	Infit MNSQ	Outfit MNSQ
<i>Seeking External Feedback Through Monitoring (SEFM)</i>				
1. I check whether I have mastered the subject content by doing extra exercises	-0.02	0.03	0.81	0.8
2. I check whether I have fully understood the subject content by doing past exam papers	-0.10	0.03	0.86	0.84
3. I keep track of my progress by recording my performance	0.11	0.03	0.98	1.02
4. I ask myself questions in my head to check whether I have understood the subject content	-0.20	0.03	0.98	0.95
5. I check my performance against the answers in textbooks or on a website	0.22	0.06	1.16	1.21
<i>Seeking External Feedback Through Inquiry (SEFI)</i>				
6. I ask my teachers to give me feedback about my performance	0.25	0.03	1	1
7. I ask my family members to give me advice on my work	-0.13	0.03	1.22	1.18
8. I ask my friends to tell me how to improve my learning	-0.12	0.03	1.11	1.07
9. I ask my fellow group members to evaluate my contributions to group work tasks	0.00	0.05	0.9	0.89
<i>Seeking Internal Feedback (SIF)</i>				
10. My gut feelings** (<i>hinala o pakiramdam</i>) tell me whether my work is good or bad	-0.10	0.03	1.01	1.02
11. My emotions influence my evaluation on my learning performance	0.04	0.03	0.99	0.98
12. How my body feels (e.g. feeling tired or energetic) tells me how well I am doing	0.05	0.03	0.97	0.99
13. My intuition** (<i>kutob</i>) tells me if I am doing a good job or not	0.01	0.05	1.06	1.1
<i>Self-reflection (SR)</i>				
14. I seek out the reasons for mistakes I made after getting back marked work	-0.09	0.03	1.02	1
15. I think about how much sense the comments of other people (e.g. teachers, family members, and friends) regarding my work make to me	0.16	0.03	0.96	0.99
16. Any areas I am unsure of after finishing my work, I go over again	0.27	0.03	0.88	0.91
17. I think about whether the way I am studying is really helping me learn	-0.31	0.03	0.93	0.9
18. 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	-0.01	0.03	1.04	1.01
19. I pay attention to my assessment results in order to identify what I can do better next time	-0.32	0.03	1.27	1.19
20. I reflect on my weaknesses when I discuss study-related issues with my classmates	0.30	0.07	1.16	1.23

Note. *All measures and SEs are in logits. **To aid readability, this word is followed by a translation to the Philippine local language.

imputation by chained equation (Azur, Stuart, Frangakis, & Leaf, 2011). Participants with extreme scores on multiple items or variables which might bias the parameter estimates (i.e. multivariate outliers; Kline, 2015) were identified using the Mahalanobis distance rule (see Bedrick, Lapidus, & Powell, 2000). This rule evaluates how far each person's score lies from the

centroid of all the participants (i.e. the aggregated mean of all item means). Fifty-seven participants were identified as outliers and were excluded from the CFA. Next, the two SaPS structural models were tested: a one-factor model and a four-factor model (SEFM, SEFI, SIF, and SR). All items were treated as continuous data, and the following goodness-of-fit indices were used to evaluate and compare the models: Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (SRMR). Following the [Hu and Bentler \(1995\)](#) recommendation, a good model fit would include model CFI and TLI of greater than 0.90 and an RMSEA of less than .08. An SRMR value of less than .08 is considered a good fit ([Hu & Bentler, 1999](#)).

A multidimensional Rasch-based model ([Adams, Wilson & Wang, 1997](#)) using ConQuest 2.0 ([Wu, Adams, Wilson & Haldane, 2007](#)) was employed in this study because self-assessment practice was conceptualized as a multidimensional construct ([Yan & Brown, 2017](#)). All subscales are calibrated simultaneously, and the measurement precision on each subscale can be enhanced by taking into account any intercorrelations between the subscales ([Bond, Yan, & Heene, 2020](#)). The indicators used to examine the instrument quality in Rasch analysis include Rasch reliability, response category functioning, and item MNSQ (mean squared) fit statistics (i.e. infit MNSQ and outfit MNSQ).

For the between-network validity, we used bivariate correlations and structural equation modeling (SEM), respectively. Specifically, we evaluated how each of the hypothesized SaPS subscales correlate to agentic, cognitive, and metacognitive engagement followed by SEM. The use of a full SEM is essential to further examine the association between the constructs while controlling for item-level measurement errors ([Yu & Hsu, 2013](#); [Zumbo, 2014](#)). We first tested whether the three engagement outcomes would fit a three-factor model. Then, we regressed the agentic, cognitive, and metacognitive engagement outcomes to the four SaPS factors. The same data screening procedures with the SaPS CFA were applied to the engagement outcomes. Consequently, the engagement data ($n = 706$) were merged to the SaPS data ($n = 702$), resulting in a final analytic dataset ($n = 673$) for the SEM. Apart from the Rasch analysis, all other analyses were conducted through the statistics software R ([R Core Team, 2016](#)) using [Rosseel's \(2012\)](#) lavaan package.

Results

Within-Network Construct Validity

CFA results showed that the one-factor model did not have good model fit (see [Table 3](#)), whereas the four-factor model had good model fit: $\chi^2(164) = 444.47$, CFI = .934, TLI = .923, RMSEA = .049, and SRMR = .04. The four-factor model also had significantly better fit than the one-factor model [$\chi^2_{diff}(6) = 300.16$, $p < 0.001$]. [Figure 1](#) shows the standardized factor loadings for each of the items and their corresponding factors.

The multidimensional rating scale Rasch model results support the SaPS good psychometric properties. The step calibrations, or the measures of the transition points between adjacent categories, increased monotonically from -1.31 , -0.85 , -0.77 , 0.62 , to 2.31 logits, suggesting that the six-point response scale functioned well. All items demonstrated sufficient fit to the Rasch model as the values of infit and outfit MNSQ fell within the acceptable range between 0.75 and 1.33 ([Wilson, 2005](#)). This result indicates that all items were assessing each latent trait as hypothesized. The Rasch reliabilities for SEFM, SEFI, SIF, and SR were 0.84, 0.79, 0.75, and 0.87, respectively. [Table 2](#) presents the item difficulties, standard errors, and item fit statistics.

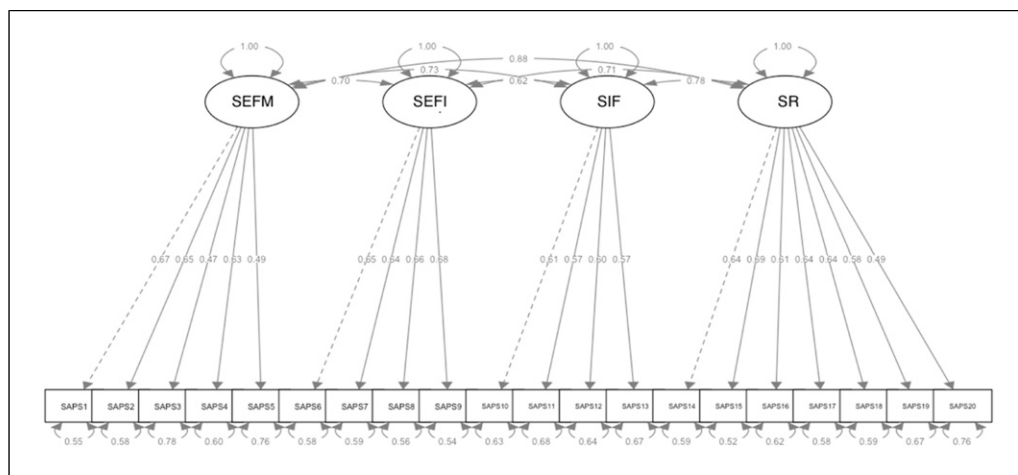
Between-Network Construct Validity

The correlation between the four SaPS factors ranged from 0.44 to 0.66, indicating a significant and moderate positive relationship across the factors. The SaPS factors were also positively

Table 3. Confirmatory factor analysis and structural equation modeling fit indices

Model	χ^2	Df	CFI	TLI	RMSEA	SRMR
CFA: SAPS 20 4-factor model ^a	444.474 ^{***}	164	0.934	0.923	0.049	0.04
CFA: SAPS 20 1-factor model ^a	744.636 ^{***}	170	0.864	0.848	0.069	0.051
CFA: Engagement 3-factor model ^b	468.825 ^{***}	86	0.927	0.911	0.079	0.048
SEM: SaPS predicting engagement ^c	1483.013 ^{***}	602	0.911	0.902	0.047	0.045

Note. CFA: confirmatory factor analysis; CFI: comparative fit index; RMSEA: root mean square error of approximation; SaPS: self-assessment practice scale; SEM: structural equation modeling; SRMR: standardized root mean square residual; TLI: Tucker-Lewis index. * $p < .05$., ** $p < .01$., *** $p < .001$; superscripts indicate sample size: a = 702, b = 706, c = 673.

**Figure 1.** The 4-factor structure of Self-assessment Practice Scale.

correlated with the engagement outcomes (see Table 1). The SEM results show that each SaPS factor predicts unique association in each engagement outcome (see Table 4; Figure 2). SEFI ($B = 0.26, p < 0.001$) and SR ($B = 0.63, p < 0.001$) were both associated with agentic engagement. Self-reflection was associated with cognitive engagement ($B = 0.88, p < 0.001$). SIF ($B = 0.28, p < 0.001$) and SR ($B = 0.60, p < 0.001$) were both associated with metacognitive engagement. SEFM had no significant association with the engagement outcomes. The full model had good fit indices: $\chi^2(538) = 1352.54$, CFI = .917, TLI = .908, RMSEA = .047, and SRMR = .044.

Discussion

This study aimed to validate the SaPS on an English language subject domain within a Philippine sample. Overall, the study highlights the SaPS' reliability and validity for use in a specific subject domain. The CFA supported the hypothesized four-factor structure, and the multidimensional Rasch analysis showed that all 20 items of the SaPS demonstrated satisfactory item/subscale fit. These findings demonstrated within-network validity. As with previous validation studies of the SaPS (Yan, 2018a, 2020a; Yan, Brubacher, Boud, Powell et al., 2020), the four-factor structure had a better structural model fit. This supports the theoretical underpinnings of the SaPS as aligning with the cyclical process of self-assessment (Yan & Brown, 2017).

The relationships between the SaPS factors with the engagement outcomes demonstrated between-network validity. As self-assessment practices (Brown & Harris, 2013; Panadero,

Table 4. Structural equation model where SaPS factors were entered as predictors of engagement outcomes.

		B	SE	P	Std. Beta
Agentic engagement<--	SEFM	0.090	0.141	0.523	0.075
	SEFI	0.294	0.068	0.000	0.296
	SIF	-0.177	0.104	0.089	-0.141
	SR	0.755	0.179	0.000	0.603
Cognitive engagement<--	SEFM	0.043	0.161	0.788	0.033
	SEFI	-0.084	0.073	0.251	-0.077
	SIF	-0.051	0.116	0.658	-0.037
	SR	1.207	0.210	0.000	0.880
Metacognitive engagement<--	SEFM	0.105	0.163	0.521	0.080
	SEFI	-0.110	0.075	0.142	-0.102
	SIF	0.368	0.120	0.002	0.271
	SR	0.844	0.200	0.000	0.623

Note. SEFM: seeking external feedback through monitoring; SEFI: seeking external feedback through inquiry; SIF: seeking internal feedback; SR: self-reflection.

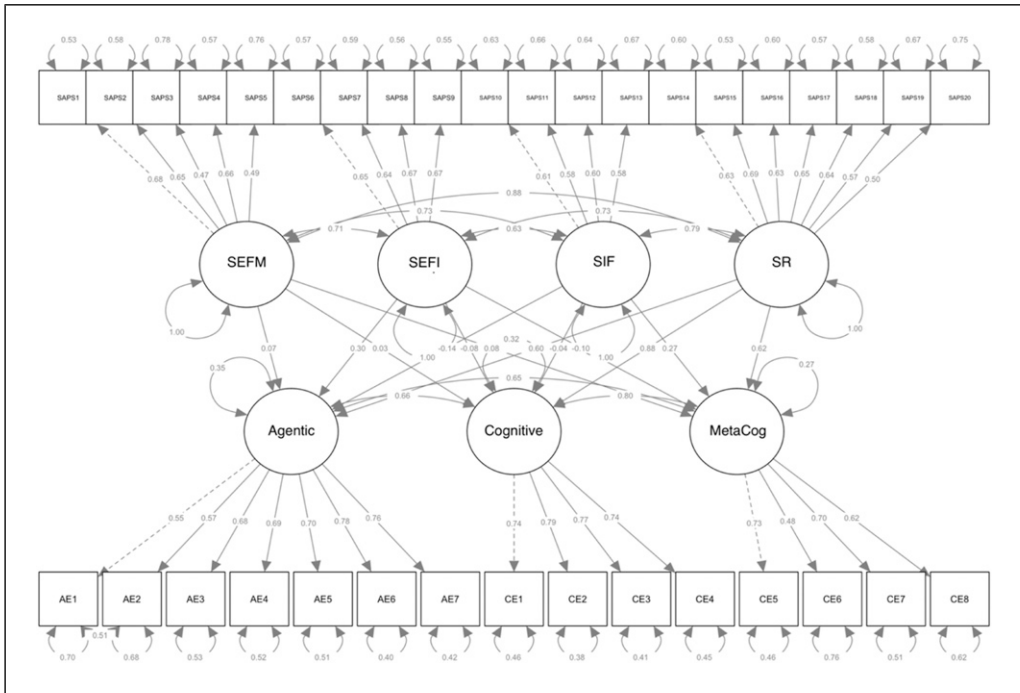


Figure 2. Structural equation model with SaPS subscales predicting agentic, cognitive, and metacognitive learning strategies.

Jonsson et al., 2016; Yan, Brown, Lee, & Qial, 2020) and student engagement (Fredricks et al., 2004; Zimmerman & Schunk, 2004) are both related with SRL, the direction of the relationship between the SaPS factors and engagement outcomes was expected to be positive. Indeed, the strength of the relationship between self-assessment and engagement was positive and moderate, indicating that the subscales are correlated but not so high as to be interpreted as tapping the same underlying construct.

Controlling for item-level measurement errors, SR predicted all engagement outcomes, especially cognitive engagement. The strong association of SR to cognitive engagement can be due to its practices similar with cognitive engagement (e.g., relating previous to new knowledge; connecting learning to personal experiences; Fredricks & McColskey, 2012; Wolters, 2004). SEFI was primarily associated with agentic engagement due to the similarity of students' intentional desire to seek feedback (see Reeve & Tseng, 2011). Both SEFI and agentic engagement involves the act of asking questions to members of their learning environment (Reeve & Tseng, 2011; Yan, 2018a; Yan & Brown, 2017). The link between SIF and metacognitive engagement can be due to the importance of internal/psychological mechanisms relevant to self-assessment and engagement (Wolters, 2004; Yan, 2018a). How students feel emotionally in school can influence their engagement (Pekrun & Linnenbrink-Garcia, 2012) and their self-assessment practices (Yan, 2018a). SEFM was not associated with engagement outcomes. This lack of association can be due to the practical use of SEFM. Students may pay more attention to self-assessment practices while learning in school (e.g., inquiry, internal feedback, and self-reflection) and less attention to self-monitoring practices (e.g., checking content mastery by re-doing past exam papers, reviewing test results against answers in the textbook or website). SEFM involves practices which are more applicable before or after school. Hence, SEFM practices may be less salient when the latter self-assessment practices are practically more applicable. The statistically significant and theoretical association of the SaPS factors to the engagement outcomes further supports the scale's validity.

The validation of the subject-specific SaPS has two potential practical implications. First, the SaPS' scores can be used to identify student self-assessment practices that need further improvement. Interventions can then be designed to target specific self-assessment practices. Second, the effectiveness of scalable interventions like monitoring logs (Zimmerman & Kitsantas, 1997), self-assessment diaries (Yan, Chiu, & Ko, 2020), or self-assessment checklists (see Meusen-Beekman et al., 2016 for self-assessment strategies used in randomized controlled interventions) can be evaluated using the subject-specific SaPS. Consequently, teachers can encourage students to seek feedback and reflect on their English learning tasks.

Limitations and Directions for Future Research

Despite the notable strengths of the present study, we also note some limitations. First, the study used cross-sectional data; hence, future studies can explore test–retest reliability and predictive validity using a longitudinal approach. Second, our sample is only composed of students from Grades 7 to 10, limiting the possible generalizability of our findings. Extending the current sample to students from primary and higher education would be a meaningful future research endeavor. Finally, consistent with previous studies is the finding that shows a relatively lower reliability of SIF. Yan and Brown (2017) showed that internal feedback is more salient for subjects involving performance-related activities (e.g., sports, music, and arts) and could be less noticeable for academic or less performance-oriented subjects like English. Hence, future investigations can (a) refine the items under SIF and make them more suitable for less performance-oriented subjects, or (b) identify alternative ways, rather than self-reports, to capture students' practices in SIF for self-assessment. From a substantive and theoretical standpoint, the exploration of social and psychological predictors of self-assessment practice are meaningful future research directions.

Conclusion

Schools from countries with an educational reform that only recently incorporated formative assessment strategies in the curriculum and geared at improving its students' performance on international assessments (i.e. PISA) could benefit by using instruments that gauge how students engage in self-assessment practices. Students' use of self-assessment practices is an invaluable 21st-

century skill that students can use for lifelong learning. This study supports the psychometric properties of the original SaPS and extends its applicability across cultures and in an English language subject domain.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Norman B. Mendoza  <https://orcid.org/0000-0003-0344-0709>

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

Note

1. The Philippines did not fare well in its first participation in the recent PISA 2018 assessments (OECD, 2019); ranking lowest on Reading achievement. Evidence suggests that self-assessment practice through self-assessment diaries can improve academic achievement (Yan, Chiu, & Ko, 2020). Studies also highlight that self-assessment practice is key for lifelong and deep learning (e.g., Boud, 1995; Papanthymou, 2018; Yan & Brown, 2017); hence, validating this scale to study, evaluate, and improve student self-assessment is an important step in improving both short- and long-term achievement.

References

- Adams, R. J., Wilson, M., & Wang, W.-c. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, *21*(1), 1-23. doi:[10.1177/0146621697211001](https://doi.org/10.1177/0146621697211001)
- Andrade, H. L. (2019). A critical review of research on student self-assessment [Systematic Review]. *Frontiers in Education*, *4*(87), 1-13. doi:[10.3389/feduc.2019.00087](https://doi.org/10.3389/feduc.2019.00087)
- Andrade, H. L., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice*, *48*(1), 12-19. doi:[10.1080/00405840802577544](https://doi.org/10.1080/00405840802577544)
- Andrade, H. L., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, *27*(2), 3-13. doi:[10.1111/j.1745-3992.2008.00118.x](https://doi.org/10.1111/j.1745-3992.2008.00118.x)
- Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: What is it and how does it work?. *International Journal of Methods in Psychiatric Research*, *20*(1), 40-49. doi:[10.1002/mp.329](https://doi.org/10.1002/mp.329)
- Bedrick, E. J., Lapidus, J., & Powell, J. F. (2000). Estimating the mahalanobis distance from mixed continuous and discrete data. *Biometrics*, *56*(2), 394-401.
- Bond, T. G., Yan, Z., & Heene, M. (2020). *Applying the rasch model: Fundamental measurement in the human sciences*. New York, NY: Routledge.
- Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.
- Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment*, (pp. 367-395). Thousand Oaks, CA: SAGE Publications, Inc.
- Chang, M.-L., & Engelhard, G. (2016). Examining the teachers' sense of efficacy scale at the item level with rasch measurement model. *Journal of Psychoeducational Assessment*, *34*(2), 177-191. doi:[10.1177/0734282915593835](https://doi.org/10.1177/0734282915593835)
- Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: A comparative analysis of various methods and student self-report instruments. In *Handbook of research on student engagement*. (pp. 763-782). New York, NY: Springer Science + Business Media.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59-109. doi:[10.3102/00346543074001059](https://doi.org/10.3102/00346543074001059)

- Hu, L. T., & Bentler, P. M. (1995). Evaluating model fit. In *Structural equation modeling: Concepts, issues, and applications*, (pp. 76-99). Thousand Oaks, CA: Sage Publications, Inc.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. doi:10.1080/10705519909540118
- Kline, R. B. (2015). *Principles and practice of structural equation modeling*. New York, NY: The Guilford Press.
- Lent, R. W., Brown, S. D., & Gore Jr, P. A. (1997). Discriminant and predictive validity of academic self-concept, academic self-efficacy, and mathematics-specific self-efficacy. *Journal of Counseling Psychology*, 44(3), 307-315. doi:10.1037/0022-0167.44.3.307
- Martin, A. J. (2007). Examining a multidimensional model of student motivation and engagement using a construct validation approach. *British Journal of Educational Psychology*, 77(2), 413-440.
- Martin, A. J., & Marsh, H. W. (2006). Academic resilience and its psychological and educational correlates: A construct validity approach. *Psychology in the Schools*, 43(3), 267-281. doi:10.1002/pits.20149
- McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice*, 10(2), 209-220. doi:10.1080/0969594032000121289
- McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons*, 87(1), 40-49.
- Meusen-Beekman, K. D., Joosten-ten Brinke, D., & Boshuizen, H. P. A. (2016). Effects of formative assessments to develop self-regulation among sixth grade students: Results from a randomized controlled intervention. *Studies in Educational Evaluation*, 51, 126-136.
- OECD. (2019). Pisa 2018 insights and interpretations. <https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf>
- 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., Brown, G. T. L., & Strijbos, J.-W. (2016). The future of student self-assessment: A review of known unknowns and potential directions [journal article]. *Educational Psychology Review*, 28(4), 803-830. doi:10.1007/s10648-015-9350-2
- Panadero, E., Jonsson, A., & Strijbos, J.-W. (2016). Scaffolding self-regulated learning through self-assessment and peer assessment: Guidelines for classroom implementation. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation*, (pp. 311-326). New York, NY: Springer International Publishing.
- 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. doi:10.1016/j.edurev.2017.08.004
- Papantymou, A. (2018). Student self-assessment in higher education: The international experience and the greek example. *World Journal of Education*, 8(6), 130-146. doi:10.5430/wje.v8n6p130
- Paris, S. G., & Newman, R. S. (1990). Developmental aspects of self-regulated learning. *Educational Psychologist*, 25(1), 87-102. doi:10.1207/s15326985ep2501_7
- Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In *Handbook of research on student engagement*, (pp. 259-282). Boston, MA: Springer.
- Pintrich, P. R., Smith, D. A. F., Garcia, T., & Mckeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (mslq). *Educational and Psychological Measurement*, 53(3), 801-813. doi:10.1177/0013164493053003024
- R Core Team. (2016). R: A language and environment for statistical computing. <https://www.R-project.org/>.
- Reeve, J., & Tseng, C.-M. (2011). Agency as a fourth aspect of students' engagement during learning activities. *Contemporary Educational Psychology*, 36(4), 257-267.
- Reeve, J., Cheon, S. H., & Jang, H. (2020). How and why students make academic progress: Reconceptualizing the student engagement construct to increase its explanatory power. *Contemporary Educational Psychology*, 62, 101899.
- Rosseel, Y. (2012). Lavaan: An r package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36. doi:10.18637/jss.v048.i02

- Siegesmund, A. (2017). Using self-assessment to develop metacognition and self-regulated learners. *FEMS Microbiology Letters*, 364(11), 1-4. doi:[10.1093/femsle/fnx096](https://doi.org/10.1093/femsle/fnx096)
- Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics*. MA, USA: Pearson Boston, Vol. 5.
- Testa, I., Capasso, G., Colantonio, A., Galano, S., Marzoli, I., Scotti di Uccio, U., ... Zappia, A. (2019). Development and validation of a university students' progression in learning quantum mechanics through exploratory factor analysis and rasch analysis. *International Journal of Science Education*, 41(3), 388-417. doi:[10.1080/09500693.2018.1556414](https://doi.org/10.1080/09500693.2018.1556414)
- West, C., Baker, A., Ehrich, J. F., Woodcock, S., Bokosmaty, S., Howard, S. J., & Eady, M. J. (2020). Teacher disposition scale (tds): Construction and psychometric validation. *Journal of Further and Higher Education*, 44(2), 185-200. doi:[10.1080/0309877X.2018.1527022](https://doi.org/10.1080/0309877X.2018.1527022)
- Wigfield, A., Eccles, J. S., Mac Iver, D., Reuman, D. A., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology*, 27(4), 552-565. doi:[10.1037/0012-1649.27.4.552](https://doi.org/10.1037/0012-1649.27.4.552)
- Wigfield, A., Guthrie, J. T., Tonks, S., & Perencevich, K. C. (2004). Children's motivation for reading; domain specificity and instructional influences. *The Journal of Educational Research*, 97(6), 299-309. doi:[10.3200/JOER.97.6.299-310](https://doi.org/10.3200/JOER.97.6.299-310)
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum Associates.
- Wolters, C. A. (2004). Advancing achievement goal theory: Using goal structures and goal orientations to predict students' motivation, cognition, and achievement. *Journal of Educational Psychology*, 96(2), 236-250. doi:[10.1037/0022-0663.96.2.236](https://doi.org/10.1037/0022-0663.96.2.236)
- Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *Acer conquest, version 2.0: Generalized item response modelling software*. Camberwell, Victoria: Australian Council for Educational Research.
- 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). 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](https://doi.org/10.1007/s40299-018-0371-8)
- Yan, Z. (2018b). Student self-assessment practices: The role of gender, school level and goal orientation. *Assessment in Education: Principles, Policy & Practice*, 25(2), 183-199. doi:[10.1080/0969594x.2016.1218324](https://doi.org/10.1080/0969594x.2016.1218324)
- Yan, Z. (2020a). Developing a short form of the self-assessment practices scale: Psychometric evidence. *Frontiers in Education*, 4, 153. doi:[10.3389/feduc.2019.00153](https://doi.org/10.3389/feduc.2019.00153)
- Yan, Z. (2020b). Self-assessment in the process of self-regulated learning and its relationship with academic achievement. *Assessment & Evaluation in Higher Education*, 45(2), 224-238. doi:[10.1080/02602938.2019.1629390](https://doi.org/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](https://doi.org/10.1080/02602938.2016.1260091)
- Yan, Z., Brown, G. T. L., Lee, J. C.-K., & Qiu, X.-L. (2020). Student self-assessment: Why do they do it? *Educational Psychology*, 40(4), 509-532. doi:[10.1080/01443410.2019.1672038](https://doi.org/10.1080/01443410.2019.1672038)
- Yan, Z., Brubacher, S., Boud, D., & Powell, M. (2020). Psychometric properties of the self-assessment practice scale (SaPS) for professional training contexts: Evidence from confirmatory factor analysis and rasch analysis. *International Journal of Training and Development*, 24(4), 357-373. doi:[10.1111/ijtd.12201](https://doi.org/10.1111/ijtd.12201)
- Yan, Z., Chiu, M. M., & Ko, P. Y. (2020). Effects of self-assessment diaries on academic achievement, self-regulation, and motivation. *Assessment in Education: Principles, Policy & Practice*, 25(5), 562-583. doi:[10.1080/0969594X.2020.1827221](https://doi.org/10.1080/0969594X.2020.1827221)
- Yu, S.-C., & Hsu, W.-H. (2013). Applying structural equation modeling methodology to test validation: An example of cyberspace positive psychology scale. *Quality & Quantity*, 47(6), 3423-3434. doi:[10.1007/s11135-012-9730-3](https://doi.org/10.1007/s11135-012-9730-3)

- Zimmerman, B. J., & Kitsantas, A. (1997). Developmental phases in self-regulation: Shifting from process goals to outcome goals. *Journal of Educational Psychology*, *89*(1), 29-36. doi:[10.1037/0022-0663.89.1.29](https://doi.org/10.1037/0022-0663.89.1.29)
- Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 299-315). New York, NY: Routledge/Taylor & Francis Group.
- Zimmerman, B. J., & Schunk, D. H. (2004). Self-regulating intellectual processes and outcomes: A social cognitive perspective. In *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development*, (pp. 323-349). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Zumbo, B. D. (2014). Structural equation modeling and test validation. In N. N. Balakrishnan, T. Colton, B. Everitt, W. Piegorisch, F. Ruggeri, & J. L. Teugels (Eds.), *Wiley statsref: Statistics reference online*. Chichester: John Wiley & Sons, Ltd.