Assessment in Education: Principles, Policy & Practice

ISSN: 0969-594X (Print) 1465-329X (Online) Journal homepage: www.tandfonline.com/journals/caie20

Teacher instructional practices and strategies for student self-assessment

Serafina Pastore, Zi Yan & Hongling Lao

To cite this article: Serafina Pastore, Zi Yan & Hongling Lao (04 Jun 2025): Teacher instructional practices and strategies for student self-assessment, Assessment in Education: Principles, Policy & Practice, DOI: 10.1080/0969594X.2025.2510204

To link to this article: https://doi.org/10.1080/0969594X.2025.2510204

ASSESSMENT IN EDUCATION: PRINCIPLES, POLICY & PRACTICE
https://doi.org/10.1080/0969594X.2025.2510204

Teacher instructional practices and strategies for student self-assessment

Serafina Pastore [a], Zi Yan [b] and Hongling Lao [b]

[a]Department of Research and Humanities Innovation, University of Bari A. Moro, Bari, Italy; [b]Department of Curriculum and Instruction, The Education University of Hong Kong, Hong Kong SAR, China

ABSTRACT

As a fundamental component of formative assessment, student self-assessment is considered *a sine qua non condition* for autonomous learners. Despite the extensive body of literature regarding principles and practice guidelines for student self-assessment, scant is the evidence about what instructional practices teachers use in classroom to promote student self-assessment. This paper reports on a research study aimed to analyse how teachers approach practices and strategies to actively encourage students to self-assess their learning. A questionnaire on self-assessment instructional practices was administered to a sample of primary- and middle-school teachers in Italy. Drawing on data collected from 246 teachers, we found that teachers tend to use different practices to promote student' self-assessment, although they resort more frequently to self-reflection strategies. This study represents a useful step in understanding teacher assessment practice and thinking about education paths for teacher assessment literacy.

ARTICLE HISTORY
Received 9 July 2022
Accepted 18 May 2025

KEYWORDS
Self-assessment; teaching practice; teacher assessment literacy; teacher professional development

Introduction

According to the landmark definition proposed by Black and Wiliam, formative assessment consists of a set of activities carried out by teachers and students to collect information 'to be used as feedback to modify teaching and learning' (1998: 140). Rather than focusing on what students have achieved, formative assessment helps to identify learning gaps, scaffold new learning, anticipate future teaching steps, and promote students' self-regulation of learning. From this perspective, self-assessment corresponds to the process through which students seek feedback on their performance, reflect on their learning, assess its quality against defined criteria, identify their learning needs, and modify learning strategies (Allal, 2020; Andrade & Brookhart, 2020; Yan, 2022; Harris & Brown, 2018).

Formative self-assessment has been widely recognised as a student key competence for effective (self-regulated) learning (Andersson & Palm, 2017; Andrade & Brookhart, 2020; Brandmo et al., 2020; Panadero et al., 2016). As an alternative assessment strategy, it involves students more actively in their own learning and, when suitably organised, it can

CONTACT Serafina Pastore serafina.pastore@uniba.it Department of Research and Humanities Innovation, University of Bari A. Moro, Bari, Italy
Supplemental data for this article can be accessed online at https://doi.org/10.1080/0969594X.2025.2510204

lead to learning enhancements (Yan et al., 2020; Pekrun, 2020). The evidence that self-assessment represents a powerful lever for the improvement of learning outcomes has been steadily accumulating over the years. During the last decades, indeed, substantial reviews of research have documented the impact of self-assessment on student learning (Yan et al., 2022, 2023; Brown & Harris, 2013; Panadero et al., 2016). However, when the focus is on the knowledge and skills a teacher needs to enact assessment in the classroom (assessment literacy), there are few insights into what self-assessment instructional practices and strategies teachers use to foster students' involvement and promote their self-regulation of learning (Panadero et al., 2014). As it is possible to develop students' self-assessment capacity through well-designed instruction (Yan & Brown, 2017; Yan, 2022), it is important to know teachers' strategies and practices in promoting student self-assessment.

Current conceptualisations of teacher assessment literacy recall the need for a solid grounding in formative assessment (Brookhart, 2023; DeLuca et al., 2020; Pastore, 2023). Self-assessment, as a component of formative assessment, should be acknowledged and used by teachers. Recognising that teachers' instruction and support are relevant to developing student self-assessment (especially for younger students), in this study, we focus on the strategies teachers use in the classroom to foster student self-assessment.

On the backdrop of the literature on teacher assessment literacy, formative assessment, and self-assessment, we report on an explorative study purposed to analyse teachers' strategies and practices for student self-assessment. Finally, we critically reflect on the main implications for research and practice in the educational assessment field.

Conceptual framework

The literature on teacher assessment literacy, formative assessment, and self-assessment lays out the foundation for the framework adopted in this study.

Self-assessment as a component of teacher (formative) assessment literacy

Teacher assessment literacy, namely what a teacher should know and be able to do with assessment, represents a research topic with a long history in the educational field (Plake & Impara, 1997; Schafer, 1993; Stiggins, 1991). The first definition attempts can be traced back over the last 65 years, when assessment literacy was identified with standardised measurement and classroom testing (Roeder, 1972). Provocatively recalled by Stiggins (1991, 1995) and Popham (1991) as a lever to restrain the US educational crisis, this concept, over the years, has been paralleled by the debate on teacher professional standards and teacher education in the assessment domain (Pastore, 2023; Van Orman et al., 2024). The widespread recognition of the socio-cultural and contextual nature of assessment (Willis et al., 2013), along with the high political and research commitment to formative assessment (Black & Wiliam, 1998; 2018, Wiliam & Black, 1996), has led to new conceptualisations of assessment literacy, currently conceived as a distinctive trait of teacher professionalism (Oo et al., 2021; Boström & Palm, 2023). Encompassing methods and strategies that provide information on student learning to be used by teachers and/or by students to enhance learning, achievement, and academic self-regulation (Andrade & Heritage, 2018; Black & Wiliam, 1998, 2018; Scott, 2006; Wiliam & Black, 1996),

formative assessment has been progressively recognised as crucial for teacher assessment literacy (Brookhart, 2023; DeLuca & Bellara, 2013; Will et al., 2019; Stiggins, 2017). In addition to attempts to conceptualise formative assessment, its components, strategies, and instruments, an extensive body of models, guidelines, and principles has been produced to support teachers in effectively embedding formative assessment into their daily practice (Adie et al., 2020; Heritage, 2013). However, despite the considerable efforts, both in the educational assessment and teacher education communities (Boström & Palm, 2020; Lyon et al., 2020), teachers continue to demonstrate substantial levels of assessment illiteracy, also in formative assessment. Albeit teachers recognise the importance of using formative assessment, they struggle to implement formative assessment and practice its main strategies. Formative assessment takes time and professional support (Egert et al., 2020; Guskey et al., 2014; Li et al., 2023); teachers struggle to integrate it into their daily routine (Wylie & Lyon, 2020) and tend to perceive this assessment as very demanding in terms of professional responsibilities (Cisterna & Gotwals, 2018). Some studies, indeed, have pointed out that formative assessment is seldom observed in practice (Bell et al., 2008; Gotch et al., 2021; Wylie & Lyon, 2015), or that its implementation is far less than satisfactory (Yan & Brown, 2021). Research evidence, furthermore, addressed how teachers challenge providing feedback to students (Gamlem, 2015; Harris & Brown, 2013). Other studies, instead, have addressed how teacher formative assessment practices are often misaligned with educational policies (e.g. professional standards) or with educational assessment literature (DeLuca et al., 2012; Klinger et al., 2012; Liang et al., 2020; Walland & Darlington, 2021). Moreover, it has already been demonstrated that teachers are reluctant to change their assessment practices (and their conceptions), especially when new practices are framed within the rationale of institutional reforms. In the Italian school system, for example, the Ministry of Education, at the end of 2020, launched a new reform that provides formative assessment as a compulsory practice for primary school teachers.

Assessment literate teachers are expected to use formative assessment and, therefore, also self-assessment. In a holistic perspective of teacher (formative) assessment literacy (Pastore & Andrade, 2019; Pastore & Mincu, 2024), self-assessment elements (e.g. concepts, methods, strategies, contextual, and social aspects) are considered entangled with teacher identity and teacher assessment practices in the classroom. Assuming that (formative) assessment literate teachers can differentiate assessment methods and strategies, the present study attempted to understand whether or not (and to what extent) primary school teachers in Italy include self-assessment strategies in their formative assessment practice.

Formative self-assessment: features and conditions

Since the recognition of assessment as an integral part of instruction and learning, the need for robust inferences about student achievement has sparked an increasing interest in formative assessment, as well as in the strategies a teacher can use in support of student motivation and learning (Andrade et al., 2019; Wiliam, 2017). While the first studies attempted to define formative assessment features (Heritage, 2010; Wiliam & Thompson, 2008) the research, over the last decades, has concentrated on the extent to which teachers realise formative assessment, and

to what extent formative assessment and its strategies affect students' learning outcomes (Andersson & Palm, 2017; Schildkamp et al., 2020). The attention turned to the integration of formative assessment into the process of teaching and learning (Brandmo et al., 2020; Heritage, 2013) led to identify different strategies and practices through which to collect, evaluate, and use evidence of student learning (Andrade & Brookhart, 2020; Evans & Marion, 2024). Echoing Wiliam and Thompson (2008), self-assessment is a core formative assessment strategy as it activates students as the owners of their learning (Andersson & Palm, 2017; Andrade, 2010). Within formative assessment strategies and practices, indeed, self-assessment encompasses different 'mechanisms and techniques through which students describe (i.e. assess) and possibly assign merit or worth (i.e. evaluate) the quality of their own learning processes and products' (Panadero et al., 2016, p. 804).

Educational research showed conditions and elements for an effective self-assessment (Yan et al., 2020, 2023; Panadero et al., 2014; Pekrun, 2020). Teachers can promote self-assessment by providing students with assessment criteria (Andrade & Heritage, 2018); sharing learning goals (e.g. in a rubric) (Heritage, 2013; Panadero et al., 2014); posing a set of questions purposed to instigate students to reflect on their learning (Harris & Brown, 2018); training and practising self-assessment (Gamlem, 2015; Schildkamp et al., 2020). It is supposed that these practices are correctly and frequently realised by a teacher (Brown & Harris, 2013). However, attention to teachers' literacy about self-assessment techniques and strategies, as well as attention to their professional needs related to the promotion of student self-assessment is lacking. Compared with studies investigating the effects of self-assessment on student learning, research on self-assessment practice on the teacher side remains limited. If, on the one hand, teachers are more conscious of the need for using several assessment information and, therefore, different assessment methods and strategies, on the other hand, core practices in student assessment often remain nebulous and problematic, and teachers struggle using research to change and/or improve assessment practices (Rogers et al., 2020). In addition, teachers with different backgrounds develop their own pedagogical habits and expertise (Borgmeier et al., 2016). It is worthwhile to examine group differences in teachers' self-assessment instructional practices with regard to their professional experience, content area specialisation, and school grade.

The present study

In this study, we investigate three main research questions:

(1) What are the Italian teachers' instructional practices to promote students' self-assessment in the classroom context?
(2) How do teachers' self-assessment instructional practices differ in their professional experience, content area specialisation, and school grade?
(3) Are there any meaningful clusters of teachers in terms of self-assessment instructional practices?

We expect these questions to be particularly relevant to the teacher formative assessment literacy and teacher professional development debate.

The Italian school context

The Ministry of Education, in December 2020, launched a new reform and introduced formative assessment as a compulsory practice for primary school teachers. Therefore, following the Ministerial Decree n. 172 and its related guidelines, teachers are expected to use formative assessment in classroom engaging themselves with substantiated pedagogical changes for high intellectual demands in their daily practice. More specifically, teachers are called to use assessment (*in itinere* and *ex-post*) providing students with a descriptive judgement on their learning. This descriptive judgement, articulated on four levels (advanced, intermediate, basic, and initial level of development) must consider the following aspects: the student autonomy; the learning progression; the transfer of student learning; and the self-regulation of learning. Although some of the main aspects that generally frame formative assessment are reported (Allal, 2020; Andrade & Heritage, 2018), the alignment with the rationale of this policy innovation was, from the very beginning, challenging for teachers due to the enduring COVID-19 pandemic situation (i.e. remote instruction), and the request for a substantial, for some aspects radical, change of their assessment practice during the on-going school year (2019–2020).

A massive, broad professional development programme has been designed and provided by the Ministry of Education to support teachers in the implementation of this new assessment practice. In the first phase, the training programme was provided to teachers and school principals online, through two-hour lessons and seminars aimed at explaining the rationale behind the Ministerial Decree, the main changes in the assessment practice, and the modalities to formally align with the new policy requirements. However, despite the first positive enthusiasm for the reform, teachers started to show some misconceptions about formative assessment and to demonstrate scepticism for the national professional development path designed to align assessment practice with the new educational policy requirements. More specifically, teachers complained about the lack of effective explanations on modalities through which to implement formative assessment, the difficulties in introducing this new assessment during the school year, and the lack of time to understand how to shift from the summative assessment mechanics, extensively used for years, to the new required strategies like self-assessment.

Methods

Instruments

With a focus on self-assessment instructional practices and strategies, in this study we used the *Self-assessment Practice Scale* (SaPS). The SaPS, originally developed and validated in the school context (Yan, 2018, 2020), has been successfully applied in higher education and domain-specific context (Mendoza & Yan, 2021; Yan, 2022). This scale is

made up of 20 items grouped in four sub-scales: *Seeking External Feedback Through Monitoring* and *Through Inquiry*; *Seeking Internal Feedback* and *Self-reflection*.

The four sub-scales cover the main common and sequenced actions performed by students in the self-assessment process.

More specifically:

- the scale *Seeking External Feedback Through Monitoring* (SEFM), with five items, includes practices and strategies aimed to allow students monitoring their learning and using external evidence against legitimate standards (e.g. past tests, or reference books);
- the scale *Seeking External Feedback Through Inquiry* (SEFI) includes four items focused on practices and strategies aimed allowing students to monitor their learning against subjective evidence from other people (e.g. teachers, pairs, and parents);
- the scale *Seeking Internal Feedback* (SIF) focuses on internal feedback. These 4 items highlight aspects such as emotions, feelings, and subjective internal states that can affect students learning experiences and assessment;
- the scale *Self-reflection* (SR) comprises 7 items related to the actions by which students go back and reflect upon their learning performance and outcomes.

For the aims of this study, the SaPS items have been modified considering the teachers' standpoint. More specifically, self-assessment components are proposed in terms of actions, strategies, and practices presented or suggested by the teacher to the students. To explain the procedure used to change the items from the student perspective to the teacher one, in the following we report the adaptations made to the first items of the four scales. For example: the first item of the Seeking External Feedback Through Monitoring (SEFM) scale, '*I check whether I have mastered the course content by doing extra exercises*' became '*I suggest students to check whether they have mastered the course content by doing extra exercises*'. In the same way, the item of the Seeking External Feedback Through Inquiry (SEFI) scale, '*I ask my teachers to give me feedback about my performance*' was changed into the item '*I suggest students to seek teachers' feedback about their performance*'. Also, the items of the Seeking Internal Feedback (SIF) scale have followed the same procedure. The item '*My gut feelings tell me whether my work is good or bad*' became '*I recommend students to consider their gut feelings in order to understand if their work is good or bad*'. Finally, for the Self-reflection (SR) scale, the original item '*I seek out the reasons for mistakes I made after getting back marked work*' was changed into '*I help students to seek out the reasons for mistakes they made after getting back marked work*'.

Given its focus on teachers' instructional practices to promote student self-assessment, the new version of the scale has been titled Teacher Self-Assessment Instructional Practices (TSaIPS).

A 6-point Likert-type response scale (1 = Strongly disagree; 6 = Strongly agree) was adopted for all items. The scale has been developed in Italian and piloted with 50 voluntary teacher students enrolled in the third year of a teacher education course and through a consequent peer-review. This process has ensured that the questionnaire was acceptable and understandable for the Italian context. Following the forward and backward translation procedure, the scale has been then translated into English (see the *Appendix*).

Table 1. Participants' demographic description.

Variable	N	Ratio	Variable	N	Ratio
Age			*Grade school*		
20–30	16	6.5%	Primary school	173	70.3%
31–40	45	18.3%	Middle school	73	29.7%
41–50	67	27.2%			
51–60	89	36.2%			
61+	29	11.8%			
Gender			*Content Area*		
Male	28	11.4%	Liberal arts	181	73.6%
Female	218	88.6%	STEM	65	26.4%
Teaching experience in years					
1–5	75	30.5%			
6–10	29	11.8%			
11–20	47	19.1%			
20+	95	38.6%			

Participants

Data were collected from a volunteer sample of primary and middle schools. In Italy, these grade schools are considered a unique cycle in the school system. However, it has to be noted that the reform, currently, concerns only teachers from grades 1 to 5.

A non-probability sample design was used (a snowball sample). The TSaIPS was administered to 290 teachers. All participants were informed of the purposes of the study and were assured of the anonymity and confidentiality of their responses. They voluntarily consented to participate in the study. Two hundred and eighty-seven questionnaires were gathered. Questionnaires returned incomplete or missing responses were excluded from further analyses. Two hundred and forty-six participants were included in the final dataset used for analysis.

Even though the sample does not perfectly mirror the national population of the school system, it shows the main characteristics of Italian teachers across the board (OECD, 2019). Most of them are female ($N = 218$; 88.6%). Teaching in Italy is still considered as a feminine profession (i.e. child-care practices).

Most of the participants (63.4%) were between the ages of 41 and 60 years; 38.6% of participants had more than 20 years of teaching experience. To compensate for the small sample size in some subjects (e.g. Social Science = 14), the subjects were combined into two broad categories. More specifically, the first cluster that includes Math and Science was named STEM. All remaining subjects (Italian, Arts, and Social Science) were grouped as Liberal Arts. In this group that takes up 73.6% were included also teachers for students with special learning needs. More details of the participants are available in Table 1.

Data analyses

The psychometric properties of the scale were examined with a multidimensional Rasch analysis (Adams et al., 1997) using ConQuest 2.0 (Wu et al., 2007). All subscales are calibrated simultaneously in the multidimensional Rasch model, and the inter-correlations between the subscales are utilised to increase the measurement precision (Bond et al., 2020). The indicators used to examine the psychometric properties include Rasch reliability, response category functioning, and

item fit statistics (i.e. Infit MNSQ and Outfit MNSQ). The Rasch-calibrated person measures were used for subsequent analyses. First, descriptive statistical analyses were conducted to answer research question 1. Then, to further investigate group differences in research question 2, linear regressions were conducted using SPSS 27. Furthermore, latent profile analysis was conducted to answer research question 3 using Mplus 8.3.

Results

Psychometric properties of the teacher self-assessment instructional practice scale

The multidimensional Rasch analysis showed that all 20 items, with only one exception, had Infit/Outfit MNSQ values falling in the acceptable range between 0.75 and 1.33 (Wilson, 2005). The result indicates that items were assessing the latent trait as theoretically hypothesised. The only exception is the Infit MNSQ of item #19 ('*I ask my students to pay attention to their assessment results to identify what they can do better next time*') (1.39). As there was only a marginal misfit and the item represents an important aspect of self-reflection strategies, this item was kept. The four subscales had satisfactory Rasch reliabilities: 0.89, 0.88, 0.83, and 0.90, respectively. The step calibrations (i.e. the measures of the transition points between adjacent categories) increased monotonically from −1.76, −0.85, −0.30, 0.70, to 2.21 logits, indicating that the six-point response scale functioned well (Linacre, 2006). The Wright map (Figure 1) displays the person abilities and item difficulties calibrated on the same metric. The four continua on the left side indicate the person measures on each of the four subscales. Persons are placed from top-to-bottom with descending levels of self-assessment instructional practice. On the right side, the items are grouped into the four subscales and positioned with the most challenging items at the top and the least challenging items at the bottom. The map shows that the TSaIPS provided a reasonably targeted measurement of teachers' self-assessment instructional practices, although the range of item difficulty was relatively smaller than the range of person abilities.

RQ1: exploring self-assessment practices using descriptive analyses

The correlation matrix among variables is shown in Table 2. There are positive and significant correlations among the four teacher self-assessment instructional practices. There is a positive correlation between gender and grade ($r = .38$), reflecting a higher ratio of female teachers in primary school than in middle school.

The detailed descriptive results are shown in Table 3. It indicated that among the four types of self-assessment practices, the Italian teachers taught students about 'self-reflection' the most often ($m = 1.81$), 'seeking external feedback through monitoring' next ($m = 1.62$), 'seeking external feedback through inquiry' less ($m = 1.25$), and 'seeking internal feedback' the least ($m = 0.88$). The mean values of these four self-assessment practices are shown in Figure 2.

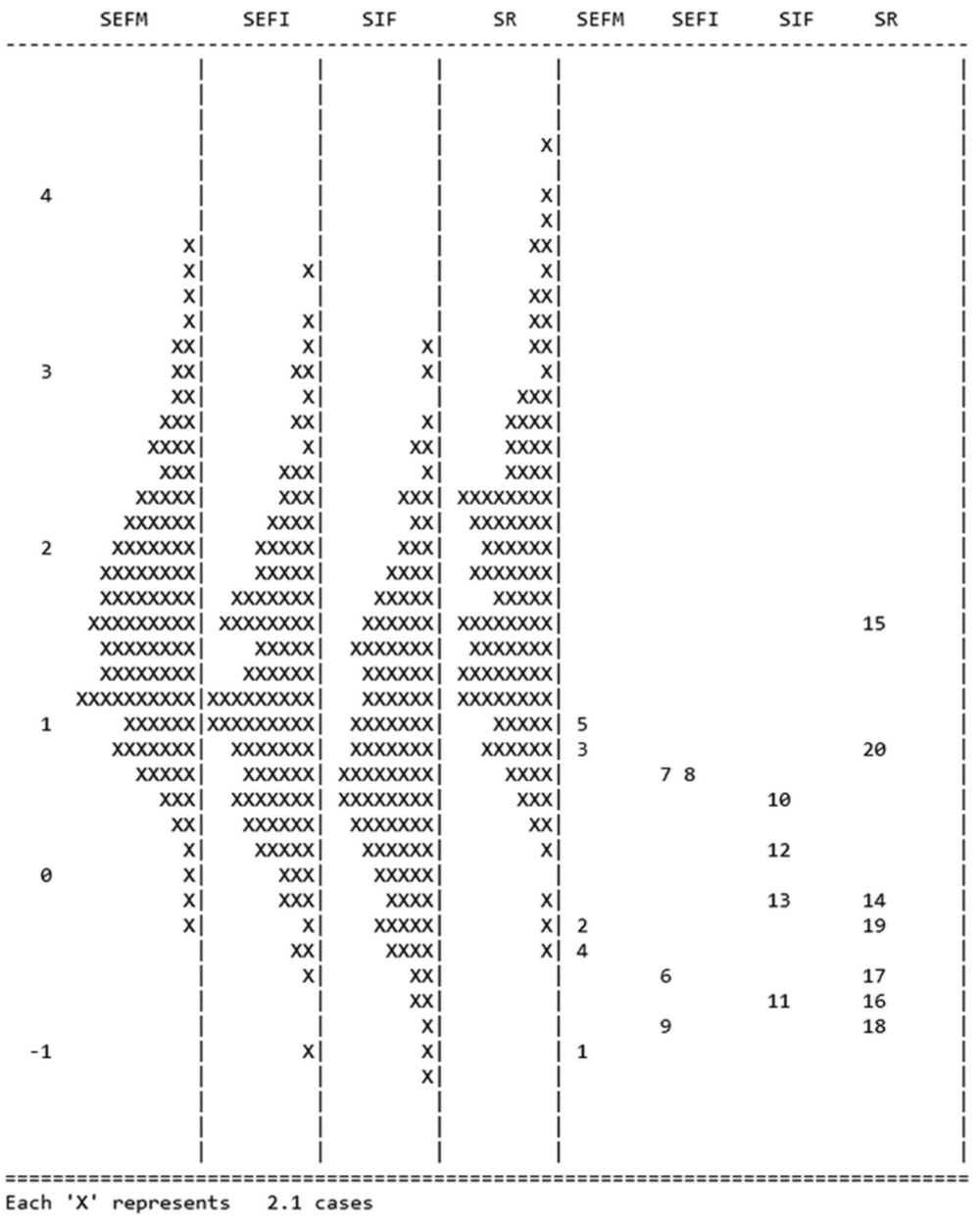

Figure 1. The Wright map of the TSalPS.

RQ2: detecting group difference using multiple linear regression

Furthermore, group difference was examined using multiple linear regression (Hair et al., 2010). Each of the four self-assessment practices was predicted by age, gender (female as baseline group), school grade (primary school as baseline group), and subject (liberal arts as baseline group), respectively.

Table 2. Correlation matrix among variables.

	Age	Gender	Grade	Subject	SEFM	SEFI	SIF	SR
Age	1.00	−.03	−.10	−.14*	.08	.13*	.07	.08
Gender		1.00	.38**	−.04	−.05	.02	−.10	.06
Grade			1.00	−.19**	−.07	−.05	−.14*	.10
Subject				1.00	.04	.05	.10	−.05
SEFM					1.00	.61**	.54**	.69**
SEFI						1.00	.64**	.68**
SIF							1.00	.58**
SR								1.00

* Correlation is significant at the 0.05 level (2-tailed); ** Correlation is significant at the 0.01 level (2-tailed); SEFM stands for 'seeking external feedback through monitoring', SEFI stands for 'seeking external feedback through inquiry', SIF stands for 'seeking internal feedback', and SR stands for 'self-reflection'.

Table 3. Mean and standard deviation of teachers' self-assessment instructional practices across gender, grade, and subject.

		N	SEFM		SEFI		SIF		SR	
Overall		246	1.62	(1.05)	1.25	(1.13)	.88	(1.12)	1.81	(1.20)
Gender	Female	218	1.64	(1.02)	1.25	(1.09)	.92	(1.09)	1.78	(1.15)
	Male	28	1.49	(1.31)	1.30	(1.43)	.59	(1.31)	2.01	(1.54)
Grade	Primary	173	1.67	(1.02)	1.29	(1.08)	.98	(1.01)	1.73	(1.11)
	Middle	73	1.51	(1.13)	1.16	(1.23)	.65	(1.32)	1.99	(1.38)
Subject	Liberal	181	1.64	(1.07)	1.28	(1.13)	.94	(1.10)	1.76	(1.14)
	STEM	65	1.57	(1.01)	1.18	(1.13)	.73	(1.17)	1.93	(1.36)

SEFM stands for 'seeking external feedback through monitoring', SEFI stands for 'seeking external feedback through inquiry', SIF stands for 'seeking internal feedback', and SR stands for 'self-reflection'. Standard deviations are enclosed in parentheses.

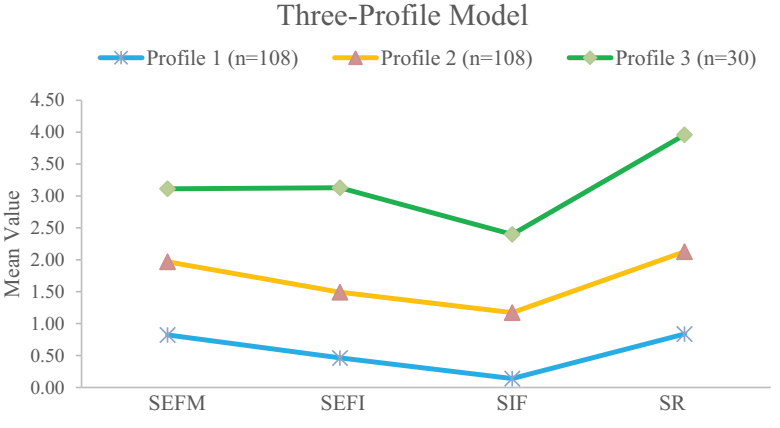

Figure 2. Line graph comparing three profiles on teachers' self-assessment instructional practices.

$$y = b_0 + b_1 * Age + b_2 * Gender + b_3 * Grade + b_4 * Subject$$

Results of four sets of regressions are shown in Table 4. Except for age on SEFI, none of the other coefficients was statistically significant. In general, with an increase in age, teachers took more frequent application of teacher self-assessment instructional practices, with a minimum difference in SEFM (b = 0.008) and a maximum difference in SEFI (b = 0.15*). In other words, senior teachers used SEFI more. As for the other three

Table 4. Results of multiple linear regressions on teachers' self-assessment instructional practices.

Predictor \ Coefficient DV	SEFM		SEFI		SIF		SR	
	b	s.e.	b	s.e.	b	s.e.	b	s.e.
Age	.008	.007	**.015***	.007	.007	.007	.010	.007
Gender	−.082	.230	.139	.244	−.191	.242	.100	.260
Grade	−.104	.164	−.111	.174	−.231	.173	.249	.186
Subject	−.072	.158	−.133	.168	−.18	.166	.076	.179

Female was the baseline group for Gender, primary school was the baseline group for Grade, liberal arts was the baseline group for Subject. * indicates statistically significant at .05 level. For SEFM, F(4,241) = 0.73, $p = 0.58$. For SEFI, F(4,241) = 1.44, $p = 0.22$. For SIF, F(4,241) = 1.77, $p = 0.14$. For SR, F(4,241) = 1.27, $p = 0.14$.

predictors, it showed a consistent pattern. Although not significant, teachers from baseline groups used SEFM, SEFI, and SIF slightly more than those from compared groups, whereas the opposite was true for SR. For example, compared with primary school teachers (baseline group), the mean score of secondary school teachers was 0.104 lower on SEFM (b = −0.104, s.e. = 0.164), and 0.249 higher on SR (b = 0.249, s.e. = 1.86). A similar pattern was found for gender and subject.

RQ3: identifying meaningful clusters using latent profile analysis

With regard to research question 3, a latent profile analysis was conducted (Nylund et al., 2007; Spurk et al., 2020). To decide on the number of profiles, a series of models with one to five profiles were fit and compared. The model fit indexes were summarised in Table 5. First, the Lo-Mendell-Rubin adjusted LRT tests clearly indicated a preference for the three-profile model than the four-profile model ($p = .168$), indicating no statistically significant improvement in model fit after adding an additional profile, although five-profile model has the highest certainty in classification (entropy = .88). Secondly, according to the three criteria information indexes (AIC, BIC, ABIC), there was a big improvement from two-profile to three-profile model, but less so from three-profile to four-profile. Furthermore, the smallest profile for the three-profile model took up 12%, while that for the four-profile and five-profile models was 3%, less than 5%, which indicated an unstable categorisation. It might be due to a small sample size or a true small proportion in this category. In addition, compared to the three-profile model, there is no new pattern in the four-profile model, adding little extra value to theoretical understanding. Thus, a parsimonious three-profile model was selected.

A description of the three profiles is shown in Table 6, with a visual representation shown in Figure 2. Profile 1 has 108 teachers (44%), with low values in all four teacher self-assessment instructional practices (SEFM = .85, SEFI = .47, SIF = .18, and SR = .87). Teachers in this group hardly seek internal feedback. Profile 2 has 108 teachers (44%), with medium values in all four teacher self-assessment instructional

Table 5. LPA model fit summary.

No. of profiles	LL	FP	AIC	BIC	ABIC	LMR-LRT (p)	BLRT (p)	Entropy	Smallest profile
1	−1509.46	8	3034.09	3062.14	3036.78	–	–	–	
2	−1357.93	13	2741.85	2787.42	2746.21	.011	0	.85	.23
3	*-1283.10*	*18*	*2602.21*	*2665.31*	*2608.25*	*.010*	*0*	*.84*	*.12*
4	−1263.59	23	2573.17	2653.80	2580.89	.168	0	.84	.03
5	−1247.82	28	2551.63	2649.78	2561.02	.211	0	.88	.03

LL stands for log likelihood; FP stands for the number of free parameters.

Table 6. Three-profile model results.

		Profile 1 Low in SA (n = 108)		Profile 2 Medium in SA (n = 108)		Profile 3 High in SA (n = 30)	
		Mean	*s.e.*	*Mean*	*s.e.*	*Mean*	*s.e.*
SEFM		.82	(.07)	1.97	(.10)	3.11	(.24)
SEFI		.46	(.08)	1.49	(.09)	3.13	(.36)
SIF		.14	(.09)	1.17	(.10)	2.40	(.23)
SR		.84	(.07)	2.13	(.10)	3.96	(.24)
Age		46.90	(10.84)	50.28	(9.99)	49.93	(10.02)
		N	*Ratio*	*N*	*Ratio*	*N*	*Ratio*
Grade	Primary	70	64.8%	86	79.6%	17	56.7%
	Middle	38	35.2%	22	20.4%	13	43.3%
Subject	Liberal	76	70.4%	87	80.6%	18	60.0%
	STEM	32	29.6%	21	19.4%	12	40.0%

Table 7. Covariate analysis result for the three-profile model.

Variable	Profile 1 Low in SA (*n* = 108)		Profile 2 Medium in SA (*n* = 108)	
Age	.97	(.02)	1.00	(.02)
Grade	.76	(.40)	.45*	(.27)
Subject	.80	(.39)	.49	(.26)

Profile 3 (High in SA, $n = 30$) served as the reference group. The value is odd ratio, standard error estimates are enclosed in parentheses. * $p < .05$.

practices (SEFM = 1.98, SEFI = 1.51, SIF = 1.18, and SR = 2.17). Teachers from this group are more active in using self-assessment instructional practices, compared to the first group, most frequently using 'self-reflection' and 'seeking external feedback via monitoring'. Profile 3 has 30 teachers (12%), with the most active self-assessment instructional practices (SEFM = 3.15, SEFI = 3.17, SIF = 2.42, and SR = 3.95). Moreover, the demographic information for three profiles was added in the same table, for reference. The average age for Profile 1 was 46.90 (SD = 10.84), the youngest of three groups. The average age for Profile 2 was 50.28 (SD = 9.99), the oldest of the three groups. The average age for Profile 3 was 49.93 (SD = 10.02), quite close to that of Profile 2. For school grade distribution, Profile 2 has the highest ratio of primary school teachers ($n = 86$, $r = 79.6\%$), whereas Profile 3 has the highest ratio of secondary school teachers ($n = 17$, $r = 43.3\%$). For subject distribution, Profile 1 and Profile 2 were similar, with the majority from liberal arts ($r = 70.4\%$, 80.6%, respectively), whereas teachers from Profile 3 had a relatively balanced subject background ($r = 60\%$, 40% for Liberal arts and STEM respectively).

To further investigate for potential group differences in profile membership, a three-step covariate analysis was conducted in Mplus, using Age, Grade and Subject as predictors. Profile 3 was treated as the reference group. Results in Table 7 showed a statistically significant difference in profile membership between grades (odd ratio = .45, s.e. = .26, $p = .03$). It meant the odds of membership in Profile 2 relative to Profile 3 was .45 times lower for teachers from middle school compared to those from primary school. In other words, teachers from primary schools were more likely to be classified in Profile 2 rather than in Profile 3, compared to teachers from middle school, and vice versa.

Discussion

A core aspect of formative assessment is to involve students in being responsible for their own learning (Black & Wiliam, 2018; Heritage, 2013). Among the different strategies to collect, interpret, and use evidence of student learning to both adjust teaching and learning towards learning goals (Wiliam & Thompson, 2008), student self-assessment is crucial (Andrade, 2019). Teachers, in fact, are not the sole source of information on student learning in the classroom. Evaluating their own learning processes and products (Panadero et al., 2016), students can generate useful insights for improving their learning. From this perspective, educational research has widely pointed out the positive impact of self-assessment on student performance and self-regulation of learning (Yan, 2022; Brown & Harris, 2013; Panadero et al., 2014). Students need to build and develop their own strategies to self-regulate their learning process. Consequently, teachers should assist their students by providing them with 'opportunities to update their understandings of concepts and skills and revise and improve their performance' (Andrade & Heritage, 2018, p. 88). Assuming teachers can encourage students to self-assess their learning, in this section, we summarise the findings from this study, with a focus both on understanding current teachers' instructional practices and strategies for promoting student self-assessment and on possible avenues for improving this practice moving forward.

First of all, the Rasch analysis showed that the instrument used in this study, i.e. the Teacher Self-assessment Instructional Practices Scale, had satisfactory psychometric properties. Based on the data collected via the instrument, we found that even though the introduction of formative assessment is relatively new in the Italian school system, teachers in this study demonstrate the use of all four types of self-assessment instructional practices (RQ1). The most frequent self-assessment instructional strategies are related to the idea of promoting student self-reflection (SR); followed by seeking external feedback via monitoring (SEFM) and via inquiry (SEFI); the less used strategies, instead, pertain to internal feedback (SIF). Not only does SIF have the lowest scores of all factors, but it is also the least endorsed by middle school teachers.

A cultural explanation can be found for this result. It is likely that these teachers tend to suggest self-assessment more in terms of self-reflection and monitoring rather than in terms of seeking internal feedback. It is likely that these teachers, when inviting students to self-assess their learning, prioritise self-reflection and monitoring, which are objectively based criteria for making a self-assessment. Socio-emotional aspects of the self-assessment practice (SIF), related, for example, to the personal awareness of feelings, or the idea of integrating and comparing personal feedback with the feedback received by family members, are less considered by teachers. While internal feedback is recognised by the research literature and by the self-assessment model used for this study, data show how teachers have a somewhat partial view of self-assessment components and strategies, more focused on student performances (especially middle school teachers). An aspect that is confirmed by the multivariate analysis of covariance performed.

The comparison of teachers' self-assessment instructional practices and strategies for variables such as gender, age, and subject matter revealed no significant differences, except for age on SEFI (RQ2). In other words, older teachers used SEFI more frequently than younger teachers. From this perspective, senior teachers tend to use strategies to promote self-assessment in their students focusing more on the revision/improvement of

students' own work/performance. Therefore, senior teachers prompt students searching for different forms of feedback inquiring with people such as peers or parents.

The latent profile analysis, finally, allowed a clustering of teachers in terms of self-assessment practices (RQ3). This analysis, indeed, shows how few teachers ($N = 30$) are more active in promoting self-assessment in their students. While the internal feedback strategies (e. g, *I suggest to students to consider their intuition in telling them if they are doing a good job or not*) are hardly proposed to students, especially by teachers of Profile 1 and 2, the self-reflection strategies represent the most used by teachers (in all three profiles). Such findings may imply a potential learning progression on self-assessment skills, with self-reflection strategies as the most intuitive and internal feedback strategies as most estranged to this population. Accordingly, it may be more effective to teach self-reflection strategies first and internal feedback strategies last while designing instructional materials on self-assessment skills, either to students or to teachers. Furthermore, the group difference identified via the latent profile analysis could be utilised to promote self-assessment practices. For example, teachers in Profile 3 who are most active in self-assessment instructional practices are valuable resources for knowledge transfer to those in the other two profiles who are less active/knowledgeable in self-assessment practices. A learning community could be formed with teachers from different profiles sharing practices and strategies and fostering teachers' formative assessment literacy (Gamlem, 2015; Poortman et al., 2022).

Results show a somewhat irregular image of teachers' self-assessment instructional practices. This does not imply that our study participants do not use the strategies defined in the four dimensions self-assessment model used for the study. However, given the recent educational policy emphasis on formative self-assessment in the Italian school system and the lack of national research studies on this topic, it is likely that the teachers involved in this study do not have a clear understanding of what self-assessment is and how it should be taught to students differentiating strategies and practices.

This aspect, considered the new emphasis on formative assessment, as framed by the last reform in the Italian school system, represents a criticality and highlights the difficulties teachers have in understanding institutional reforms and in transferring instructional innovations (i.e. formative assessment and self-assessment) into the daily classroom routine (Walland & Darlington, 2021).

The implications in terms of teacher education and professional development are different. While educational research identified the features of effective teacher professional development programs on formative assessment (e.g. active learning, teacher collaboration, professional learning community, appropriate duration of time spent on the programme) (Boström & Palm, 2020; Poortman et al., 2022), no sufficient evidence has been provided, in the Italian school system, to support teachers to align with the new policy requirements.

Our results point out the importance of understanding teacher practices used to foster student self-assessment practice. Moreover, it has to be noted the gap between the sophisticated models of student self-assessment (e.g. the model used for this study with its four dimensions of self-assessment) and the feeble interest of educational research in investigating how teachers teach their students to be responsible for tracking their learning progress. The gathered evidence of the positive effects of teachers' instructional

practices of self-assessment is currently limited, within the formative assessment framework, to sharing success criteria or co-creating success criteria and monitoring learners' progress through their self-assessment (Andrade & Brookhart, 2020). In this perspective, the teacher self-assessment instructional practices are related only to the dimensions of seeking external feedback through monitoring and through inquiry.

Several concerns arise here in terms of the realistic use of data for teachers' (and students') decision-making. It has been argued (Andrade & Brookhart, 2020; Harris & Brown, 2018) that interventions to promote student self-assessment should consider enhancing the clarity of learning goals or involving students in reflecting on the final product of learning outcomes. However, educational research (Harris & Brown, 2018) has rarely explored if and how teachers support their students in practising self-assessment skills and offer them feedback on these skills. Moreover, despite the consensus on the positive benefits for students, the documentation of the student self-assessment implementation, as well as the teachers' tensions and difficulties in the use of self-assessment practice are still inchoative (Harris & Brown, 2018; Panadero et al., 2016). If self-assessment, as one of the main components of formative assessment, is part of the concept of teacher assessment literacy, the implications for pre- and in-service education are relevant. In this perspective, for example, it should be interesting to understand if, and how, the dimension of seeking internal feedback in self-assessment practice is related to the socio-emotional aspects or to the teacher image as assessor currently advocated in the conceptualisations of teacher assessment literacy (Pastore & Andrade, 2019; Looney et al., 2017; Xu & Brown, 2016). Additional investigations into teacher practices and strategies oriented to student self-assessment, as well as into teacher development paths, are now warranted.

Limitations

These findings have to be interpreted in the light of some limitations. First, our data came from a small sample; therefore, the findings cannot be generalised to the entire Italian school system. Second, our data are self-reported data.

Future research should explore factors that might account for the variability in teachers' self-assessment strategies. Factors to consider include teacher demographic characteristics, differences across disciplines, and differences in teacher training or professional development. There was not sufficient information and/or variability in our study to allow us to detect the role these factors may play in teacher self-assessment instructional practices. A key point in the investigation on teachers' (formative) assessment literacy and self-assessment practices is represented by professional development (as well as by teacher education programmes) that should prepare teachers with the knowledge, mindset, attitudes, insights, commitment and confidence necessary to consistently build assessment competence and address different assessment practices.

Despite its limitations, however, the study presents a unique opportunity to address challenges, in theory and practice, to review teacher education and professional development in the assessment domain. Future research exploring self-assessment instructional practices of teachers would benefit from qualitative studies (e.g. through observations) to determine whether and how the use of self-assessment practices in the classroom matters most.

Conclusion

Despite the widespread recognition and the growing interest in student self-assessment, the instructional practices and strategies used by teachers promoting student self-assessment, in this study, appear somewhat scattered and not effectively embedded in the teaching-learning process. The different use of strategies in the four main dimensions of the SaPS (Yan, 2018) demonstrates that self-assessment instructional practices of teachers involved in this study seem still far from the perspective of an assessment really embedded in a learning culture.

The improvement of teacher (formative) assessment literacy represents one of the most pressing and contested contemporary educational policy and practice issues (Brookhart, 2023; Van Orman et al., 2024). Despite its inner limitations, the present study confirms how challenging it can be for teachers to implement self-assessment in the classroom (Gamlem, 2015; Poortman et al., 2022). The attention on the strategies teachers adopt to encourage student self-assessment suggests a careful consideration of teacher assessment literacy, and consequently of the modalities through which teachers acquire, nurture, and develop their assessment literacy and data-use culture. The implications for the development of assessment literate teachers are several if we advocate self-assessment not only as a set of tools but as a crucial element in assessment practice (e.g. teacher decision-making; student metacognitively informed improvement of learning outcomes). Future studies should strive to better understand how teachers orient and support their students in the self-assessment practice.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Serafina Pastore is an Associate Professor in the Department Research and Humanities Innovation, University of Bari A. Moro, Bari, Italy. Her research examines the complex intersections of assessment practice, teacher education and educational policy as operating within the context of school and university innovations. Her recent work focuses on teacher assessment literacy.

Zi Yan is a Full Professor in The Education University of Hong Kong, Hong Kong SAR, China. His publications and research interests focus on two related areas, i.e., educational assessment in the school and higher education contexts with an emphasis on student self-assessment; and Rasch measurement, in particular its application in educational and psychological research. A recent book co-edited with Lan Yang is entitled, *Assessment as learning: Maximising opportunities for student learning and achievement.*

Hongling Lao is a Post-doctoral Fellow in the Department of Curriculum and Instruction at The Education University of Hong Kong, Hong Kong SAR, China. She majored in educational measurement at graduate school, specialized in diagnostic classification models. Her current research interest is the integration of educational measurement and assessment to enhance learning and teaching in K-12 context, with a focus on teacher and student assessment literacy.

ORCID

Serafina Pastore http://orcid.org/0000-0001-8598-021X
Zi Yan http://orcid.org/0000-0001-9305-884X
Hongling Lao http://orcid.org/0000-0001-9417-8333

References

Adams, R. J., Wilson, M., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, *21*(1), 1–23. https://doi.org/10.1177/0146621697211001

Adie, L., Stobart, G., & Cumming, J. (2020). The construction of the teacher as expert assessor. *Asia-Pacific Journal of Teacher Education*, *48*(4), 436–453. https://doi.org/10.1080/1359866X.2019.1633623

Allal, L. (2020). Assessment and the co-regulation of learning in the classroom. *Assessment in Education Principles, Policy & Practice*, *27*(4), 332–349. https://doi.org/10.1080/0969594X.2019.1609411

Andersson, C., & Palm, T. (2017). The impact of formative assessment on student achievement: A study of the effects of changes to classroom practice after a comprehensive professional development programme. *Learning & Instruction*, *49*, 92–102. https://doi.org/10.1016/j.learninstruc.2016.12.006

Andrade, H. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In H. Andrade & G. Cizek (Eds.), *Handbook of formative assessment* (pp. 90–105). Routledge.

Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, *4*. https://doi.org/10.3389/feduc.2019.00087

Andrade, H. L., Bennett, R. E., & Cizek, G. J. (Eds). (2019). *Handbook of formative assessment in the disciplines*. Routledge. https://doi.org/10.4324/9781315166933

Andrade, H. L., & Brookhart, S. M. (2020). Classroom assessment as the co-regulation of learning. *Assessment in Education Principles, Policy & Practice*, *27*(4), 350–372. https://doi.org/10.1080/0969594X.2019.1571992

Andrade, H. L., & Heritage, M. (2018). *Using formative assessment to enhance learning, achievement, and academic self-regulation*. Routledge.

Bell, C., Steinberg, J., Wiliam, D., & Wylie, C. (2008). *Formative assessment and teacher achievement: Two years of implementation of the keeping learning on track program*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York, NY.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice Policy & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education Principles, Policy & Practice*, *25*(6), 1–25. https://doi.org/10.1080/0969594X.2018.1441807

Bond, T. G., Yan, Z., & Heene, M. (2020). *Applying the Rasch Model: Fundamental measurement in the human sciences*. Routledge.

Borgmeier, C., Loman, S. L., & Hara, M. (2016). Teacher self-assessment of evidence-based classroom practices: Preliminary findings across primary, intermediate and secondary level teachers. *Teacher Development*, *20*(1), 40–56. https://doi.org/10.1080/13664530.2015.1105863

Boström, E., & Palm, T. (2020). Expectancy-value theory as an explanatory theory for the effect of professional development programmes in formative assessment on teacher practice. *Teacher Development*, *24*(4), 539–558. https://doi.org/10.1080/13664530.2020.1782975

Boström, E., & Palm, T. (2023). The effect of a formative assessment practice on student achievement in mathematics. *Frontiers in Education*, *8*(), 1101192. https://doi.org/10.3389/feduc.2023.1101192

Brandmo, C., Panadero, E., & Hopfenbeck, T. N. (2020). Bridging classroom assessment and self-regulated learning. *Assessment in Education Principles, Policy & Practice*, *27*(4), 319–331. https://doi.org/10.1080/0969594X.2020.1803589

Brookhart, S. M. (2023). Assessment literacy in a better assessment future. *Chinese Journal of Applied Linguistics*, *46*(2), 162–179. https://doi.org/10.1515/CJAL-2023-0202

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). Sage.

Cisterna, D., & Gotwals, A. M. (2018). Enactment of ongoing formative assessment: Challenges and opportunities for professional development and practice. *Journal of Science Teacher Education*, *29*(3), 200–222. https://doi.org/10.1080/1046560X.2018.1432227

DeLuca, C., & Bellara, A. (2013). The current state of assessment education: Aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education*, *64*(4), 356–372. https://doi.org/10.1177/0022487113488144

DeLuca, C., Schneider, C., Coombs, A., Pozas, M., & Rasooli, A. (2020). A cross-cultural comparison of German and Canadian student teachers' assessment competence. *Assessment in Education Principles, Policy & Practice*, *27*(1), 26–45. https://doi.org/10.1080/0969594X.2019.1703171

Egert, F., Dederer, V., & Fukkink, R. G. (2020). The impact of in-service professional development on the quality of teacher-child interactions in early education and care: A meta-analysis. *Educational Research Review*, *29*, 100309. https://doi.org/10.1016/j.edurev.2019.100309

Evans, C., & Marion, S. (2024). *Understanding instructionally useful assessment*. Routledge.

Gamlem, S. M. (2015). Feedback to support learning: Changes in teachers' practice and beliefs. *Teacher Development*, *19*(4), 461–482. https://doi.org/10.1080/13664530.2015.1060254

Gotch, C. M., Carbonneau, K. J., & Van Orman, D. S. J. (2021). Preparing preservice teachers for assessment of, for, and as learning. In S. L. Nichols & D. Varier (Eds.), *Theory to practice: Educational psychology for teachers and teaching* (Vol. 1, pp. 17–36). Information Age Publishing.

Guskey, T. R., Roy, O., & von Frank, V. (2014). *Reaching the highest standard in professional learning*. Corwin Press & Learning. https://doi.org/10.1080/03057640500146880

Hair, J. F., Anderson, R. E., Babin, B. J., & Black, W. C. (2010). *Multivariate data analysis: A global perspective* (7th ed.). Pearson Education.

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 & Teacher Education*, *36*, 101–111. https://doi.org/10.1016/j.tate.2013.07.008

Harris, L. R., & Brown, G. T. L. (2018). *Using self-assessment to improve student learning* (1st ed.). Routledge.

Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. Corwin Press.

Heritage, M. (2013). *Formative assessment in practice: A process of inquiry and action*. Harvard Education Press.

Klinger, D. A., Volante, L., & DeLuca, C. (2012). Building teacher capacity within the evolving assessment culture in Canadian Education. *Policy Futures in Education, Special Issue: Developing Sustainable Assessment Cultures in School Learning Organisations*, *10*(4), 447–460. https://doi.org/10.2304/pfie.2012.10.4.447

Li, Z., Yan, Z., Chan, K. K. Y., Zhan, Y., & Guo, W. Y. (2023). The role of a professional development program in improving primary teachers' formative assessment literacy. *Teacher Development*, *27*(4), 447–467. https://doi.org/10.1080/13664530.2023.2223595

Liang, X., Collins, L. J., Lenhart, L., & Ressa, V. (2020). Instructional change following formative instructional practices professional development. *Teacher Development*, *24*(1), 108–125. https://doi.org/10.1080/13664530.2019.1705886

Linacre, J. (2006). *A users guide to WINSTEPS Ministep: Rasch-model computer programs*. Winsteps. https://www.winsteps.com/winman/copyright.htm

Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2017). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education Principles, Policy & Practice*, *25*(5), 442–467. https://doi.org/10.1080/0969594X.2016.1268090

Lyon, C. L., Nabors Oláh, L., & Brenneman, M. (2020). A formative assessment observation protocol to measure implementation: Evaluating the scoring inference. *Educational Assessment*, *25*(4), 288–313. https://doi.org/10.1080/10627197.2020.1766957

Mendoza, N. B., & Yan, Z. (2021). Validation of a subject-specific student self-assessment practice scale (SaPS) among secondary school students in the Philippines. *Journal of Psychoeducational Assessment*, *39*(4), 481–493. https://doi.org/10.1177/0734282921994374

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent Class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569. https://doi.org/10.1080/10705510701575396

OECD [The Organization for Economic Cooperation and Development]. (2019). *TALIS 2018 results (vol. 1): Teachers and school leaders as lifelong learners*. OECD Publishing.

Oo, C. Z., Alonzo, D., & Davison, C. (2021). Pre-service teachers' decision-making and classroom assessment practices. *Frontiers in Education*, *6*, 628100. http://dx.doi.org/10.3389/feduc.2021.628100

Panadero, E., Brown, G. T. L., & Courtney, M. G. (2014). Teachers' reasons for using self-assessment: A survey self-report of Spanish teachers. *Assessment in Education Principles, Policy & Practice*, *21*(4), 365–383. https://doi.org/10.1080/0969594X.2014.919247

Panadero, E., Brown, G. T. L., & 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. https://doi.org/10.1007/s10648-015-9350-2

Pastore, S. (2023). Teacher assessment literacy: A systematic review. *Frontiers in Education*, *8*, 1217167. https://doi.org/10.3389/feduc.2023.1217167

Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching & Teacher Education*, *84*, 128–138. https://doi.org/10.1016/j.tate.2019.05.003

Pastore, S., & Mincu, M. (2024). A feasible balance? The Italian teachers' standpoint on assessment literacy, assessment practice, and teacher professional development. *European Journal of Teacher Education*, *47*(2), 226–245. https://doi.org/10.1080/02619768.2024.233885

Pekrun, R. (2020). Self-report is indispensable to assess students' learning. *Frontline Learning Research*, *8*(3), 185–193. https://doi.org/10.14786/flr.v8i3.637

Plake, B. S., & Impara, J. C. (1997). Teacher assessment literacy: What do teachers know about assessment? In G. D. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 53–58). Academic Press.

Poortman, C. L., Brown, C., & Schildkamp, K. (2022). Professional learning networks: A conceptual model and research opportunities. *Educational Research*, *64*(1), 95–112. https://doi.org/10.1080/00131881.2021.1985398

Popham, W. J. (2011). Assessment literacy overlooked: A teacher educator's confession. *The Teacher Educator*, *46*(4), 265–273.

Roeder, H. H. (1972). Are today's teachers prepared to use tests? *Peaboby Journal of Education*, *49*, 239–240.

Rogers, A. P., Mitescu Reagan, E., & Ward, C. (2020). Preservice teacher performance assessment and novice teacher assessment literacy. *Teaching Education*, *33*(2), 175–193. https://doi.org/10.1080/10476210.2020.1840544

Schafer, W. D. (1993). Assessment literacy for teachers. *Theory into Practice*, *32*(2), 118–126. https://doi.org/10.1080/00405849309543585

Schildkamp, K., van der Kleij, F. M., Heitink, M. C., Kippers, W. B., & Veldkamp, B. V. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research*, *103*, 101602. https://doi.org/10.1016/j.ijer.2020.101602

Scott, D. (2016). *New perspectives on curriculum, learning and assessment, evaluating education: Normative systems and institutional practices*. Springer. https://doi.org/10.1007/978-3-319-22831-0_6

Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior*, *120*, 103–445. https://doi.org/10.1016/j.jvb.2020.103445

Stiggins, R. (2017). *The perfect assessment system*. ASCD.

Stiggins, R. J. (1991). Relevant classroom assessment training for teachers. *Educational Measurement Issues & Practice*, *10*(1), 7–12. https://doi.org/10.1111/j.1745-3992.1991.tb00171.x

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, *77*(3), 238–245.

Van Orman, D. S. J., Gotch, C. M., & Carbonneau, K. J. (2024). Preparing teacher candidates to assess for learning: A systematic review. *Review of Educational Research*, *95*(3), 427–463. https://doi.org/10.3102/00346543241233015

Walland, E., & Darlington, E. (2021). How do teachers respond to assessment reform? Exploring decision-making processes. *Educational Research*, *63*(1), 80–94. https://doi.org/10.1080/00131881.2020.1857653

Wiliam, D. (2017). Assessment and learning: Some reflections. *Assessment in Education Principles, Policy & Practice*, *24*(3), 394–403. https://doi.org/10.1080/0969594X.2017.1338520

Wiliam, D., & Black, P. (1996). Meanings and consequences: A basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal*, *22*(5), 537–548. https://doi.org/10.1080/0141192960220502

Wiliam, D., & Thompson, M. (2008). Integrating assessment with instruction: What will it take to make it work? In C. A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Lawrence Erlbaum Associates.

Will, K. K., McConnell, S. R., Elmquist, M., Lease, E. M., & Wackerle-Hollman, A. (2019). Meeting in the middle: Future directions for researchers to support educators' assessment literacy and data-based decision making. *Frontiers in Education*, *4*. https://doi.org/10.3389/feduc.2019.00106

Willis, J., Adie, L., & Klenowski, V. (2013). Conceptualising teachers' assessment literacies in an era of curriculum and assessment reform. *The Australian Educational Researcher*, *40*, 241–256. https://doi.org/10.1007/s13384-013-0089-9

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Lawrence Erlbaum Associates.

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *Acer conquest, version 2.0: Generalised item response modelling software*. In Australian Council for Educational Research.

Wylie, E. C., & Lyon, C. J. (2015). The fidelity of formative assessment implementation: Issues of breadth and quality. *Assessment in Education Principles, Policy & Practice*, *22*(1), 140–160. https://doi.org/10.1080/0969594X.2014.990416

Wylie, E. C., & Lyon, J. C. (2020). Developing a formative assessment protocol to support professional growth. *Educational Assessment*, *25*(4), 314–330. https://doi.org/10.1080/10627197.2020.1766956

Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching & Teacher Education*, *58*, 149–162. https://doi.org/10.1016/j.tate.2016.05.010

Yan, Z. (2018). The self-assessment practice scale (SaPS) for students: Development and psychometric studies. *Asia-Pacific Educational Researcher*, *27*(2), 123–135. https://doi.org/10.1007/s40299-018-0371-8

Yan, Z. (2020). Developing a short form of the self-assessment practices scale: Psychometric evidence. *Frontiers in Education*, *4*, 153. https://doi.org/10.3389/feduc.2019.00153

Yan, Z. (2022). *Student self-assessment as a process for learning*. Routledge.

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. https://doi.org/10.1080/02602938.2016.1260091

Yan, Z., & Brown, G. T. L. (2021). Assessment for learning in the Hong Kong assessment reform: A case of policy borrowing. *Studies in Educational Evaluation*, *68*, 100985. https://doi.org/10.1016/j.stueduc.2021.100985.

Yan, Z., Brown, G. T. L., Lee, C. K. J., & Qiu, X. L. (2020). Student self-assessment: Why do they do it?. *Educational Psychology*, *40*(4), 509–532. https://doi.org/10.1080/01443410.2019.1672038

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*, *27*(5), 562–583. https://doi.org/10.1080/0969594X.2020.1827221

Yan, Z., Lao, H., Panadero, E., Fernández-Castilla, B., Yang, L., & Yang, M. (2022). Effects of self-assessment and peer-assessment interventions on academic performance: A meta-analysis. *Educational Research Review*, *37*, 100484. https://doi.org/10.1016/j.edurev.2022.100484.

Yan, Z., Wang, X., Boud, D., & Lao, H. (2023). The effect of self-assessment on academic performance and the role of the explicitness: A meta-analysis. *Assessment & Evaluation in Higher Education*, *48*(1), 1–15. https://doi.org/10.1080/02602938.2021.2012644