Studies in Educational Evaluation 88 (2026) 101578

Contents lists available at ScienceDirect

Studies in Educational Evaluation

journal homepage: www.elsevier.com/locate/stueduc

A systematic review of feedback literacy scales: Development and validation

Ying Zhan [1], Zi Yan [2], Ye Zeng [3], Wenyun Luo [4], Zhi Hong Wan [*,5]

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

ARTICLE INFO

Keywords:
Student feedback literacy
Teacher feedback literacy
Systematic review
Scale validation

ABSTRACT

The effectiveness of feedback practices in educational settings depends critically on the feedback literacy of both students and teachers. Accurately assessing feedback literacy levels is essential for facilitating teachers and students to plan for a better feedback experience. This systematic review synthesised 17 scales measuring feedback literacy through analysing their structure and psychometric properties. Our analysis revealed a predominant focus on student feedback literacy scales in a general higher education context. Furthermore, existing scales primarily assessed capacities and dispositions, often overlooking the knowledge component. Current teacher feedback literacy scales do not fully include the key elements that teachers need to help students improve their feedback literacy. While evaluations of these scales generally indicated acceptable to good psychometric reliability and validity, opportunities for methodological improvement exist, such as using multiple reliability assessment methods and conducting Rasch analysis. To improve measurement accuracy, direct tests or more behavioural indicators should be considered in scale design.

1. Introduction

Feedback is widely recognised as a fundamental and influential element in students' effective learning (Hattie & Timperley, 2007; Mandouit & Hattie, 2023). Despite teachers' substantial effort to provide feedback, students' dissatisfaction and disengagement with feedback have been consistently reported (Deeley et al., 2019). There is growing recognition that for feedback processes to be effective, shared responsibilities between students and teachers are necessary (Nash & Winstone, 2017). These shared responsibilities require students and teachers to be equipped with adequate feedback knowledge, capacities, and dispositions to actively play their corresponding roles, thus maximising feedback impacts on learning (Carless & Winstone, 2023a). Feedback knowledge, capacities, and dispositions are framed as an important concept in feedback literacy, echoing the recent shift in the feedback paradigm from teacher-centred information transmission to a more interactive, learner-centred approach (Nieminen & Carless, 2023a).

Student feedback literacy (SFL) refers to students' knowledge, competencies, and dispositions to make sense of the received feedback and use it to enhance their learning (Carless & Boud, 2018). To support students' uptake of feedback, teacher feedback literacy (TFL) was proposed as teachers' "knowledge, expertise and dispositions to design feedback processes in ways which enable student uptake of feedback and seed the development of student feedback literacy" (Carless & Winstone, 2023a, p. 153). Whether it is SFL or TFL, it is conceptualised within a tripartite framework that includes knowledge, capacities, and dispositions necessary for students and teachers to make feedback effective for learning.

Although there is a debate on if SFL or TFL can be regarded as a psychological construct (Nieminen & Carless, 2023), some researchers have made efforts in developing SFL and TFL scales, aiming to consistently measure them across different contexts, which may transform scholars' discussions on feedback literacy into a shared discourse (Zhan, 2022). In addition, by quantitatively measuring SFL and TFL, educators and researchers can build comprehensive models that clarify the roles

* Correspondence to: Department of Curriculum & Instruction, The Education University of Hong Kong, D1–1 F-33, 10 Lo Ping Rd, Tai Po, Hong Kong, China.
E-mail address: wanzh@eduhk.hk (Z.H. Wan).
[1] https://orcid.org/0000-0002-2362-2887
[2] https://orcid.org/0000-0001-9305-884X
[3] https://orcid.org/0000-0002-9845-8428
[4] https://orcid.org/0009-0004-9961-4172
[5] https://orcid.org/0000-0002-8163-5862

https://doi.org/10.1016/j.stueduc.2026.101578
Received 21 August 2025; Received in revised form 27 January 2026; Accepted 28 January 2026
Available online 10 February 2026

and impacts of feedback literacy within the complex educational ecosystem. This, in turn, informs targeted interventions and policies aimed at optimising feedback processes and enhancing overall educational quality (; Zhang et al., 2024). While some recent reviews have examined the concept of feedback literacy and related intervention (e.g., Little et al., 2024; Nieminen & Carless, 2023), to date, no systematic review has investigated the scales to measure SFL or TFL across educational contexts.

To bridge the above-mentioned review gaps, a systematic examination of SFL and TFL scales across educational levels is warranted. By assessing the structures and components as well as the psychometric quality of the existing SFL and TFL scales, the review aimed to gain insight into the differences and similarities among them and identify the limitations of the existing scale design. Consequently, it would inform both researchers and practitioners in the scale design of feedback literacy. Three specific review questions are listed below:

- What are the structures and components of SFL and TFL scales?
- How are the reliability and validity of SFL and TFL scales ensured?
- What are the implications for feedback literacy scale design?

2. Methodology

2.1. Article searching

The article search was conducted in three databases: EBSCO, Scopus, and Web of Science. The Boolean operators were used to generate the following search syntax: AB = ("feedback literacy") AND AB = ("measure*" or "assess*" OR "instrument*" OR "scale*"). The limiters for the respective databases are full-text, peer-reviewed journal papers in English. In all, one hundred and forty-nine English papers published in peer-reviewed journals were selected after cleaning duplicates from different databases. We conducted the last search in June 2025.

2.2. Article scanning

The PRISMA flow chart (Moher et al., 2009) was used to delineate the selection process of eligible articles in this review (see Fig. 1). The studies were excluded if they did not a) relate to feedback literacy; b) focus on the validation of feedback literacy scales; c) report psychometric properties; and d) include the complete items of the survey. As a result, one hundred and thirty-two studies were ruled out, and seventeen studies were retained for the systematic review.

2.3. Article coding

Relevant data were extracted into four categories according to the review questions: basic study information, structure of scales, descriptions of scales, and psychometric properties of the scales. Initially, the coding procedure began with the extraction of fundamental information from the selected studies, including details such as the author, year of publication, country, educational level, learning context of study, participants (e.g., gender, age), sample size, and sampling method. Second, the description of scales was extracted by the scale name, number of dimensions, name of dimensions, number of items, response scale characteristics, and item pooling methods. Third, evidence of reliability and validity was drawn from each study, including the reliability (e.g., Cronbach's alpha) and validity (e.g., content, substantive, and structural validity) of the scale used. To ensure coding reliability, Author 3 and Author 4 independently coded the data, achieving an intercoder agreement above 90 %. In cases of coding discrepancies, the two coders discussed the issues with Author 1 until reaching a consensus.

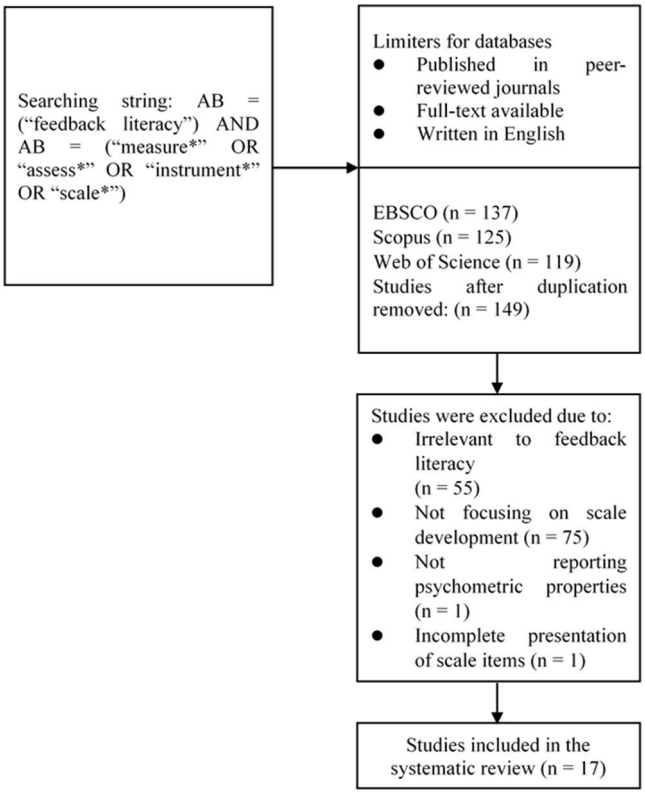

Fig. 1. Flow diagram of screening feedback literacy scales.

3. Overview of the selected articles and scales

Table 1 summarises the demographic details of seventeen selected studies. The first scale validation study was conducted by Zhan (2022). A notable peak occurred in 2023 with the development of six scales. Most studies (N = 15) were conducted in Asian countries or regions, primarily in mainland China (N = 11). Ten scales were developed for higher education, five focused on school education and two covered all the educational levels. Except for three scales developed in online or hybrid learning contexts, the others were created for conventional learning environments. Among the seventeen studies, the participants were mostly female, and more than half reported sample sizes of 600 or more. All but three studies used convenience sampling rather than random sampling.

Table 2 provides a basic description of the seventeen scales. Twelve of the scales addressed SFL. This is consistent with the observation of Nieminen & Carless (2023), who found that feedback literacy studies mostly focused on SFL. Seven studies were discipline-specific, with six focusing on writing-related contexts like academic or L2 writing. The scales included between two and six dimensions, most commonly four, and contained from ten to fifty-three items, with 76.47 % of the scales having no more than thirty items. Fourteen scales utilised literature review to develop the items, while three studies combined literature review and interview data to develop item pools. Seven scales used a 5-point Likert scale, with the degree of agreement being the most common response format. Only four scales used a frequency response scale to evaluate students' feedback behaviours.

4. Structure and components of SFL and TFL scales

The tripartite framework of feedback literacy, including knowledge, capacities and dispositions, based on the widely used conceptions of SFL (Carless & Boud, 2018) and TFL (Carless & Winstone, 2023a), was used to compare the different scales. It is worth noting that scholars used

Table 1
The demographic information of the seventeen studies.

No.	Year-Author	Country	Educational level	Learning context	Participants		Sample size	Sampling method
					Gender	Age		
P1	2022-Song	Singapore	H	O	506 males & 417 females	NA	923	CS
P2	2022-Yıldız et al.	Turkey	H	C	176 males & 559 females	NA	735	CS
P3	2022-Yu et al.	China	H	C	878 males & 1214 females	NA	2092	RS
P4	2022-Zhan	China	H	C	188 males & 367 females	NA	555	CS
P5	2023-Dong et al.	China	H	C	218 males & 256 females	17–35	474	CS
P6	2023-Lee et al.	International	M	C	170 females & 261 males	70.85 % were aged between 31 and 45	431	CS
P7	2023-Wang et al.	China	M	C	144 males & 373 females	22–59	517	CS
P8	2023-Woitt et al.	Turkey	H	C	72 males & 149 females	Mean age: 22.5 years (SD = 4.3)	221	CS
P9	2023-Yang et al.	China	S	B	188 males & 444 females	NA	632	CS
P10	2023-Zhang et al.	China	S	C	665 males & 662 females	15–23	1327	CS
P11	2024-Chen et al.	China	S	C	937 males & 1017 females	NA	1954	CS
P12	2024-Dawson et al.	International	H	C	162 males, 181 females & 7 transgender/nonbinary;	18–57	350	CS
P13	2024-Teng & Ma.	China	H	C	365 males & 343 females	19–21	708	CS
P14	2024-Zhan	China	H	C	181 males & 360 females	NA	541	CS
P15	2025-Cui et al.	China	H	O	400 males & 521 females;	642 were older than 20	921	CS
P16	2025-Kang et al.	China	S	C	NA	NA	2048	RS
P17	2025-Wancham & Tangdhanakanond	Thailand	S	C	300 males & 300 females	Grades 7–12	600	RS

Note: Publications are sorted by year from earliest to latest, and alphabetically within the same year. H=higher education; S=school education; M=mixed; C= conventional learning; O=online learning; B= hybrid learning; NA=no report; CS=convenience sampling; RS=Random sampling.

Table 2
Basic information of the seventeen scales.

No.	Scale name	Type of scale	Discipline-specific	No. of dimensions	No. of items	Likert scale	Item pooling methods
P1	Learner feedback literacy scale	S	N	3	21	5-point	LR
P2	Feedback literacy scale	S	N	4	24	5-point	LR
P3	L2 student writing feedback literacy scale	S	Y(W)	5	28	5-point	LR
P4	Student feedback literacy scale	S	N	6	24	6-point	LR
P5	Peer feedback literacy scale	S	Y(W)	4	20	6-point	LR & IR
P6	Feedback literacy scale for L2 writing teachers	T	Y(W)	3	34	5-point	LR
P7	Writing assessment feedback literacy questionnaire	T	Y(W)	4	32	5-point	LR & IR
P8	Student feedback literacy scale	S	N	2	21	5-point	LR
P9	Technology-enhanced teacher literacy scale	T	N	3	10	6-point	LR
P10	L2 secondary student writing feedback literacy scale	S	Y(W)	2	14	5-point	LR
P11	Mathematics discourse feedback skills scale	S	Y (M)	6	26	6-point	LR
P12	Feedback literacy behaviour scale	S	N	5	24	6-point	LR
P13	Metacognition-based student feedback literacy	S	Y (W)	3	53	7-point	LR
P14	Pre-service teachers' feedback literacy scale	T	N	6	28	6-point	LR & IR
P15	GenAI-Student feedback literacy scale	S	N	4	39	NA	LR
P16	A generic self- assessment scale for K-12 teachers	T	N	4	30	NA	LR
P17	Student feedback literacy test	S	N	4	25	4-point	LR

Note: T = teacher feedback literacy; S=student feedback literacy; N = no; Y=yes; W=writing; M=mathematics; NA=no report; LR=literature review; IR=interview

different terms to label the components of feedback literacy. The categorisation of components was primarily based on the classifications provided by the scale developers, as long as these classifications were explicitly stated. For components without clear classifications, we independently categorised them by analysing the items, assessing whether they reflected knowledge, capacities, or dispositions. Some components were double-coded, as shown in Table 3, because they included items pertaining to both capacities and dispositions.

4.1. SFL scales

4.1.1. Conceptual frameworks for SFL scales

The commonly used conceptual frameworks for SFL were those proposed by Carless and Boud (2018), Molloy et al. (2020), Sutton (2012), and Han and Xu (2021). Among existing scales, most studies synthesised two or more frameworks (e.g., Dawson et al., 2024; Dong et al., 2023; Zhan, 2022), others adapted and combined them with discipline- or context-specific frameworks (e.g., Chen et al., 2024; Cui et al., 2025), while a few relied on a single framework (e.g., Song, 2022; Wancham & Tangdhanakanond, 2025).

Carless and Boud's (2018) framework served as the primary foundation for most SFL scales, which describe four key characteristics of feedback-literate students: appreciating feedback, making judgments, managing affect, and taking action. To become feedback literate, students need to acquire sophisticated knowledge of feedback and positive attitudes towards it, develop their capacity to judge the quality of their work, manage their emotions effectively, and possess a repertoire of strategies to act productively. It directly underpinned the development of three scales (i.e., Wancham & Tangdhanakanond, 2025; Yu et al., 2022; Zhang et al., 2023). In addition, five scales (i.e., Chen et al., 2024; Dawson et al., 2024; Dong et al., 2023; Woitt et al., 2023; Zhan, 2022) combined Carless and Boud's (2018) framework with Molloy et al.'s (2020) empirically-driven model, a seven-dimensional framework that includes understanding feedback purposes and roles, seeking information, making judgments about work quality, managing emotions, and processing and using feedback to improve future work. For example, Dawson et al. (2024) synthesised Carless and Boud's (2018) four-component framework with the empirically derived findings of

Table 3
Dimensions of SFL scales.

No.	Knowledge	Capacities	Dispositions
P1	-	Conceptions of feedback; Feedback trust;	Self-efficacy
P2*	-		Openness to use feedback; Positive attitude; Appreciation; Awareness towards effective feedback
P3	-	Making judgement; Managing affect; Taking Actions	Appreciating feedback; Acknowledging different feedback sources
P4	-	Eliciting feedback; Processing feedback; Enacting feedback	Appreciation of feedback; Readiness to engage; Commitment to change
P5	Feedback-related knowledge	Feedback-related abilities; Cooperative learning ability	Appreciation of peer feedback; Willingness to participate
P8	-	Feedback practices	Feedback attitudes
P10 *	-	Using feedback; Evaluating feedback	Using feedback and evaluating feedback
P11	-	Comparative analysis; Expressing communication; Mathematical reasoning; Monitor and adjust; Diagnostic evaluation; Implementation capacity	-
P12	-	Seek feedback information; Use feedback information; Make sense of information; Provide feedback information; Manage affect	-
P13	Task knowledge; Strategy knowledge; Knowledge of academic writing; Knowledge of assessment	Feedback-related planning strategies; Feedback-related monitoring strategies; Feedback-related evaluating strategies; Feedback-related strategies in participation	Motivation; Confidence
P15	Feedback knowledge	Feedback prompting skills; Feedback judgement; Emotional recognition; Enactment of feedback; Feedback ethics	Feedback appreciation; Readiness to engage; Commitment to change
P17	-	Making judgments; Managing affect; Taking action	Appreciating feedback

* Note: Classifications for these scales were not specified in the original studies and were categorised by the authors.

Molloy et al. (2020), identifying concrete students' feedback behaviours.

Two scales synthesised (i.e., Cui et al., 2025; Teng & Ma, 2024) Sutton's (2012) framework, which conceptualised feedback literacy through epistemological, ontological, and practical dimensions, with other frameworks in their respective fields. For example, Cui et al. (2025) combined Sutton's (2012) three-dimensional feedback literacy framework with Ng et al.'s (2024) AI literacy model, which addresses cognitive, behavioural, affective, and ethical competencies, to develop a scale tailored to the Generative AI context. In addition, Han and Xu (2021) framed SFL as students' cognitive and socio-affective readiness to engage with feedback, which was grounded by one scale (Song, 2022). Yildiz et al. (2022) did not report a specific conceptual framework underlying their design.

4.1.2. Student feedback capacities

Student feedback capacities refer to the competencies that students need to effectively interpret, engage with, and utilise feedback to enhance their learning processes (Carless & Boud, 2018). All SFL scales included capacities, encompassing making judgments, taking actions, managing affects, eliciting feedback, and providing feedback information.

It is interesting to find that evaluative judgment - "the capability to make decisions about the quality of work of self and others" (Tai et al., 2018, p. 471) was measured by most scales. For example, Wancham and Tangdhanakanond (2025) used the term "making judgment" to characterise students' ability to evaluate and make reasonable academic decisions regarding their own and their peers' work. Chen et al. (2024) labelled evaluative judgement as "diagnostic evaluation," emphasising students' ability to assess feedback and reflect on learning strategies in a mathematical context. Additionally, through careful examination of the items, we observed that evaluative judgment was not treated as a standalone construct in some scales; instead, it was incorporated within different categories. For instance, Dong et al. (2023, p.1109) included the items related to evaluative judgement, such as "I am able to judge whether the peer comments are reasonable", under the broader components of "feedback abilities".

In contrast, students' capacity of eliciting feedback, received comparatively less attention. Students' capacity of eliciting feedback refers to students' ability to seek information from various sources to address issues related to their learning needs (Malecka et al., 2022). Students commonly employ inquiry and monitoring strategies in their feedback-seeking behaviour (Zhan, 2022). Six scales clearly addressed this capacity (Chen et al., 2024; Cui et al., 2025; Dawson et al., et al., 2024; Teng & Ma, 2024; Woitt et al., 2023; Zhan, 2022). For example, Zhan (2022) operationalised it as students' capacity to solicit information from multiple sources to improve their learning. Cui et al. (2025) situated this capacity in a Generative AI context and assessed students' ability to tailor prompts effectively in order to enhance the quality and relevance of feedback.

Students' capacity of providing feedback was least assessed in SFL scales. It was explicitly included in only one scale developed by Dawson et al. (2024, p.351), who defined it as "considering the work of others and making comments about its quality". Dong et al. (2023, p.1109) also mentioned this capacity by including relevant items such as "I am able to offer specific revision solutions to peers' writing", although it was not categorised as an independent component.

4.1.3. Student feedback dispositions

Student feedback dispositions refer to the attitudes, beliefs, and values that influence how students approach, interpret, and act upon feedback (Carless & Boud, 2018). Ten SFL scales included dispositional components, encompassing appreciation of feedback, recognition of various feedback sources, readiness to engage, commitment to change, motivation, and confidence.

Students' appreciation of feedback was commonly addressed, while students' commitment to act on feedback was least assessed. Students' appreciation for feedback was evaluated in terms of students' acknowledgement of the feedback value/usefulness/importance (e.g., Cui et al., 2025; Dong et al., 2023; Wancham & Tangdhanakanond, 2025; Woitt et al., 2023; Yıldız et al., 2022; Yu et al., 2022; Zhan, 2022; Zhang et al., 2023). For example, Cui et al. (2025, p.6) evaluated students' appreciation of feedback by discerning students' "acknowledgement of the importance and impact of GenAI-generated feedback in the learning process". Given the increasing prevalence of GenAI-generated feedback, students' acknowledgment of the values of feedback provided by GenAI has been timely captured. This also shows that SFL scales need to be adapted or developed to incorporate the new form of feedback. In terms of commitment to act on feedback, only two scales clearly assessed students' enthusiasm to engage with feedback for ongoing improvement by dedicating time and effort (Cui et al., 2025; Zhan,

2022).

4.1.4. Student feedback knowledge

Knowledge is necessary for students to develop their feedback literacy (Carless & Boud, 2018). However, the knowledge part was the least assessed by the SFL scales, which was explicitly included in only three scales (i.e., Cui et al., 2025; Dong et al., 2023, and Teng & Ma, 2024).

The existing SFL scales have been found to assess three kinds of knowledge, including feedback knowledge, domain knowledge and task knowledge. First, feedback knowledge is the knowledge that helps students engage with feedback more effectively. For example, Teng and Ma (2024) use the term "strategy knowledge" to assess students' knowledge, which enables them to participate effectively and utilise feedback within the writing process. Additionally, in Dong et al.'s (2023, p.1109) scale, there were several items, such as "I have the knowledge of judging the quality of peer feedback", used to measure the knowledge for judging the feedback. Second, domain knowledge was assessed in relation to the context of language writing. For example, Teng and Ma (2024) assessed the linguistic, genre and disciplinary knowledge. Dong et al. (2023, p.1109) also assessed it in items such as "I have the knowledge of different genres (e.g., exposition, argumentation, narration)" under the broader component of "Feedback-related knowledge". Finally, task knowledge refers to students' knowledge of the requirements of a task. Two scales (i.e., Teng & Ma, 2024; Dong et al., 2023) assessed this kind of knowledge.

4.2. TFL scales

4.2.1. Conceptual frameworks for TFL scales

The conceptual underpinnings of TFL were largely based on Carless and Winstone's (2023a) framework, which consisted of three dimensions, i.e., design, relation with students and pragmatism. Four out of five scales (i.e., Kang et al., 2025; Lee et al., 2023; Yang et al., 2023; Zhan, 2024) drew on this framework, in combination with additional conceptual (e.g., Lee et al., 2023) or empirical sources (e.g., Yang et al., 2023). For instance, Lee et al. (2023) incorporated Carless and Winstone's (2023a) framework alongside Lee's (2021) pedagogical model for L2 writing, which conceptualises teacher feedback literacy as a set of knowledge, values, and skills enacted across the stages of the feedback process. Yang et al. (2023) built on Carless and Winstone (2023a), integrating empirical findings (e.g., Espasa et al., 2019) with a focus on teacher actions in technology-enhanced feedback environments. By contrast, Wang et al. (2023) adopted Xu and Brown's (2016) assessment literacy and Taylor's (2013) language assessment literacy to capture both general feedback competencies and language-specific skills, enabling the development of a scale tailored to the Chinese EFL writing context.

4.2.2. Teacher feedback capacities

Teacher feedback capacities denote teachers' ability to design and facilitate effective feedback processes (Carless & Winstone, 2023a). These capacities were found to include teachers' abilities to plan, implement, and follow up on feedback.

Table 4 shows that most TFL scales emphasised teachers' capacities of implementing feedback. For example, Lee et al. (2023) measured "perceived skills" of handling the process of feedback. Zhan (2024) evaluated teachers' evaluative judgement, suggestion provision, and communication in the implementing stage of feedback. Similarly, Kang et al. (2025) developed a scale assessing teachers' capacity to deliver clear, relevant, constructive and supportive feedback, engage students in the feedback process and provide support for understanding and using feedback. Yang et al. (2023) in their scale focused on teachers' instructional practice of using technology to provide high-quality feedback, foster student agency during the feedback process, and sustain technology-enhanced feedback adoption using team approaches. While most TFL scales emphasised the implementation of feedback, limited attention was given to the planning and follow-up stages, which are essential for creating a positive environment for students to elicit and uptake the feedback.

Table 4
Dimensions of TFL scales.

No.	Knowledge	Capacities	Dispositions
P6	Perceived knowledge	Perceived skills	Values
P7	Knowledge of writing assessment feedback; knowledge of useful techniques	Assessment feedback competence; Assessment feedback practices	-
P9	-	The quality of feedback using technology; The support of process of feedback using technology; The practices of feedback in communities using technology	-
P14	-	Planning; Implementing; Follow-up	Appreciation; Readiness; Commitment
P16	Feedback knowledge	Feedback skills; Feedback actionability	Feedback values

4.2.3. Teacher feedback dispositions

Teacher dispositions include the attitudes and willpower to overcome challenges and to develop productive feedback processes for students (Carless & Winstone, 2023a). Three scales assessed teacher dispositions (i.e., Kang et al., 2025; Lee et al., 2023; Zhan, 2024), including their perceived values regarding feedback, readiness to communicate with students, and commitment to facilitating students' uptake of feedback.

Teachers' appreciation and perception of feedback were the foci of measurement. For example, Lee et al. (2023) included the component of "value" in their scale, which was composed of teachers' beliefs, attitudes, goals, and perceptions about feedback. Similarly, Kang et al. (2025) assessed teachers' perception of feedback, viewing it as a tool for learning rather than just evaluation. Zhan (2024) not only emphasised teachers' appreciation for feedback values but also highlighted teachers' affective and social support for students, as well as their commitment to dedicating additional time and effort to facilitate students' uptake of feedback. The underrepresentation of teachers' commitment to facilitating students' uptake of feedback, might hinder students' commitment to engaging with feedback.

4.2.4. Teacher feedback knowledge

Teacher knowledge includes teachers' understanding of feedback principles and practice (Carless & Winstone, 2023a). Three scales (i.e., Kang et al., 2025; Lee et al., 2023; Wang et al., 2023) addressed this component. Specifically, they measured teachers' perceived knowledge of feedback with less attention to domain knowledge and task knowledge. For instance, Wang et al. (2023) operationalised teachers' perceived knowledge of feedback using two distinct components: "Knowledge of Writing Assessment Feedback" and "Knowledge of Useful Techniques." The first component assessed teachers' understanding of fundamental feedback principles, such as recognising the importance of active and ongoing student engagement in the feedback process. The second component evaluated teachers' comprehension of employing diverse feedback methods and techniques to effectively support and enhance student learning outcomes. Lee et al. (2023) and Kang et al. (2025) did not classify the knowledge component but integrated these aspects into a broader construct, which encompassed familiarity with feedback principles, types, and purposes, as well as alignment with learning objectives and students' individual needs.

Table 5
The reliability of the selected scales.

No.	Internal consistency reliability					Test-retest reliability	Split-half reliability
	α	CR	ω	Inter-item correlation	Item- total correlation		
SFL							
P1	(O&D).888-.967	-	-	.357 −.887	.599 −.805	-	-
P2	(O&D).830-.940	-	-	-	.338 −.704	-	(O&D).80 −.95
P3	(D).805-.951	-	-	-	-	-	-
P4	(O&D).896-.967	-	-	-	-	-	-
P5	(D).800-.890	-	-	-	-	-	-
P8	(D).860	-	-	-	-	-	-
P10	(D).841-.923	-	-	-	-	-	-
P11	(O&D).813-.931	-	-	-	.621 −.813	-	(O).899
P12	(D).640-.810	-	-	-	-	(D).56-.71	-
P13	(O&D).820-.860	-	-	-	-	-	-
P15	(O&D).716-.948	(D).831-.874	-	-	-	-	-
P17	(O).757	-	(O) 0.698	.140 −.650	-	-	-
TFL							
P6	(O&D).800-.930	(D).808-.961	-	-	-	-	-
P7	-	(D) Above.70	-	-	-	-	-
P9	(O&D).765-.958	-	-	-	-	-	-
P14	(O&D).870-.977	-	-	-	-	-	-
P16	-	-	(O&D).89-.97	-	-	-	-

Note: CR=composite reliability; D= dimension-level; O= overall scale; O&D= overall and dimension-level.

5. Psychometric properties of SFL and TFL scales

5.1. Reliability of scales

Reliability is a key aspect of scale quality and is commonly categorised into three main types: internal consistency reliability, split-half reliability, and test-retest reliability (Hajjar, 2018). Common indicators of internal consistency include Cronbach's alpha (α), composite reliability (CR), McDonald's omega (ω), inter-item correlations, and item-total correlations. Test-retest reliability evaluates the stability of scores over time, while split-half reliability assesses internal consistency of scores between two sets.

Our analysis showed that all twelve SFL scales examined the internal consistency reliability of the scale. Cronbach's alpha was the most reported, calculated for the dimensions and/or the overall scale. Among these, eleven of twelve studies present alpha scores above 0.7, indicating the reliability ranged between acceptable and excellent. Only Dawson et al. (2024) reported questionable reliability ($0.6 <= \alpha < 0.7$) for some dimensions, such as "seeking feedback information" and "providing feedback information". Three studies conducted item-total correlation

Table 6
Validity of selected scales.

No.	No. of validity evidence	Content validity	Substantive validity	Structural validity	Generalizability validity	External validity
SFL						
P1	4	ER	RA	CFA	-	Corr among FL dimensions
P2	3	ER; LR; PR	Corr among FL dimensions	EFA and CFA	-	-
P3	4	ER; LR; PR	Corr among FL dimensions	EFA and CFA	Multi-group CFA	-
P4	5	LR	Corr among FL dimensions	CFA	Multi-group CFA	Corr between FL and other variables
P5	3	ER; LR; PR	-	EFA and CFA	-	CV: CFA factor loadings; DV: HTMT ratio; ANOVA
P8	5	ER; LR; PR	RA	EFA	DIF	CV: Corr between FL and other variables; DV: Corr between FL and other variables
P10	4	ER; PR	Corr among FL dimensions	EFA and CFA	-	SEM
P11	4	ER; LR; PR		EFA and CFA	Multi-group CFA	CV: Combined reliability and AVE; DV: Chi-square difference test
P12	5	ER	Theoretical mapping; RA	CFA	DIF	Correlation between FL and other variables
P13	4	ER; LR; PR	Corr among FL dimensions	CFA	-	SEM
P15	5	ER; LR; PR	Corr among FL dimensions	EFA and CFA	mMulti-group CFA	CV: AVE; DV: HTMT radio
P17	3	ER; LR; PR	-	CFA; IRT	Multi-group CFA	-
TFL						
P6	2	ER; LR; PR	-	EFA and CFA	-	-
P7	3	ER; LR; PR	-	EFA and CFA	-	DV: $\sqrt{AVE}$
P9	3	ER; LR; PR	-	EFA and CFA	-	CV: AVE; DV: $\sqrt{AVE}$
P14	4	ER; LR; PR	Corr among FL dimensions	EFA and CFA	-	F-tests; SEM
P16	2	ER; LR; PR	-	EFA and CFA; IRT	-	-

Note：CV= convergent validity; Corr = correlations; DIF= differential item functioning; DV= discriminant validity; ER= expert review; IRT= item response theory analysis; LR=literature review; PR= Participant review; RA= Rasch analysis; $\sqrt{AVE}$= The square root of AVE.

analyses, showing good correlation, while two reported inter-item correlations above the recommended 0.30 threshold. Only Wancham and Tangdhanakanond (2025) reported McDonald's omega (ω) with 0.698, indicating acceptable reliability. Compared with internal consistency reliability, split-half reliability was examined in two studies (Chen et al., 2024; Yıldız et al., 2022) and test–retest reliability was reported in one study (Dawson et al., 2024) with fair to good reliability.

Similar to SFL scales, the existing TFL scales all examined internal consistency reliability rather than test-retest and split-half reliability. Cronbach's alpha or McDonald's omega (ω) of four scales (i.e., Kang et al., 2025; Lee et al., 2023; Yang et al., 2023; Zhan, 2024) indicates good to excellent reliability. Lee et al. (2023) and Wang et al. (2023) reported composite reliability, ranging between acceptable and excellent. No scale reported inter-item correlations and item-total correlations.

5.2. Validity of scales

Construct validity determines whether a test measures what it claims to measure (Messick, 1995). It encompasses six key aspects: content, substantive, structural, generalizability, external, and consequential (Messick, 1995). In this review, except for consequential validity, the other five aspects were examined (see Table 6). The reason for excluding consequential validity was that no scale study has examined it during validation. Only four studies of feedback literacy addressed five aspects of validity (Dawson et al., 2024; Cui et al., 2025; Woitt et al., 2023; Zhan, 2022). The content and structural validity were tested by all scales, while the generalizability was least tested for the scales.

Among twelve scales measuring SFL, content validity was examined by all the scales through expert review (N = 11), literature review (N = 9), and participant review (N = 9). For example, Cui et al. (2025) established the theoretical framework for the scale, which is built upon a theoretical understanding of SFL and AI literacy, as outlined by Sutton (2012) and Ng et al. (2024). Several students from the target population were invited to review the items to ensure the quality, unambiguity and clarity of the items. Experts with expertise in teaching or psychometric assessment also provided some suggestions about wording adjustments.

Nine studies examined substantive validity by calculating Pearson correlations among dimensions (N = 6), Rasch analysis (N = 3), theoretical mapping (N = 1), and think-aloud tasks (N = 1). For example, Dawson et al. (2024) mentioned that substantive validity evidence in their study was ensured by correspondence between items and the underlying conceptual framework, and the item-level statistics from Rasch analysis.

To examine structural validity, six studies conducted both exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), four studies conducted CFA, one study conducted EFA, and one study conducted an item response model and CFA. For example, Yu et al. (2022) conducted an EFA to analyse the participants' responses and let the factorial structure emerge from the data. They then conducted a CFA to analyse the psychometric properties of the scales.

Nine studies examined the external validity by calculating convergent validity and discriminant validity (N = 4), Pearson correlation between feedback literacy and other variables (N = 3), structural equation modelling (N = 2) and ANOVA (N = 1). For example, to understand external validity, Dawson et al. (2024) calculated correlations between subscales of the Big Five Personality Inventory-Short Form and the feedback literacy behaviour scale.

Seven studies addressed the generalizability aspect of construct validity. Of these, five employed multi-group CFA to test measurement invariance across groups, while two studies assessed differential item functioning to examine item-level bias. For example, Zhan (2022) conducted a Multi-group CFA to test the invariance of the six-dimensional structure across sexes and majors to examine the equivalence of the SFL scale's structure for different types of participants.

Similar to SFL scales, the existing five TFL scales all examined content and structural validity, while no scale showed evidence of generalizability. Content validity was examined by expert review, participant review, and literature review on all five scales. Only Zhan (2024) examined substantive validity by calculating the correlation among the six dimensions of the TFL scale. To ensure the generalizability validity of the scales, all studies conducted EFA and CFA. Besides, Kang et al. (2025) conducted IRT analyses to demonstrate the precision of the scale in measuring feedback-giving literacy among the participants. Two studies examined external validity by calculating convergent validity and/or discriminant validity. For example, Zhan (2024) conducted an F-test and SEM to indicate the correlation between pre-service teachers' motivations for becoming a teacher and assessment course learning experience and TFL, respectively.

6. Implications

The analysis of seventeen SFL and TFL scales demonstrates the similarities and differences between them and reveals the weaknesses in the existing design, which sheds light on the future scale design of feedback literacy.

6.1. Context-specific design

It is necessary to develop context-specific feedback literacy scales for the Western context, school settings, online learning and AI settings. The review revealed that existing SFL scales were primarily developed in Asian (Eastern) contexts, where feedback is typically top-down and students are less likely to question teachers (Biggs, 1996). In Western settings, students tend to seek feedback actively and discuss it openly. In this review, the majority of the scales developed in Eastern contexts do not include students' capacities to seek feedback. The SFL scales developed in Eastern contexts may overlook or inadequately assess students' feedback behaviour, such as feedback-seeking, which is common in Western settings. Therefore, developing culture-sensitive scales that reflect these behavioral variations is essential to accurately measure feedback literacy across cultural contexts.

In addition, most feedback literacy scales have been developed in the context of higher education, with fewer in the context of school education. Schools are different from universities. For example, in Asian schools, feedback often involves whole-class activities and is influenced by a test-driven culture and large classes (Zhang et al., 2023). Therefore, students may need different feedback literacy compared to college students, who receive more personalised feedback and have less examination pressure.

The review also found that most scales have been developed for traditional face-to-face learning environments. However, with technological advancements, new feedback sources, especially AI-generated feedback, have emerged in educational contexts (Cui et al., 2025). The scales of feedback literacy in traditional learning contexts should be adapted or expanded to align with online, blended, and AI-enabled learning environments.

6.2. Discipline-specific design

Another notable design implication is the necessity of designing discipline-specific feedback literacy scales. Many of the existing scales were developed for a general learning context without setting disciplinary boundaries. However, the scales developed for a general learning context could not be fully applicable for measuring students' and teachers' discipline-specific feedback literacy. While core components of feedback literacy, such as managing emotional responses and making evaluative judgments, may be transferable across disciplines, the way feedback is interpreted and acted upon can vary significantly depending on disciplinary norms and practices. Feedback literacy needs to be contextualised and integrated with discipline-specific requirements (Malecka et al., 2022). Disciplinary tools and practices play

an important role in the feedback process, not only shaping the form and content of feedback but also influencing how participants understand and apply it (Winstone et al., 2022). For example, Teng and Ma (2024) developed metacognition-based SFL for academic writing, considering that academic writing involves a series of goal-oriented, iterative processes that necessitate students to plan, monitor, and assess their writing activities. Therefore, developing discipline-specific feedback literacy scales ensures more accurate measurement and more targeted interventions to enhance students' feedback engagement within their disciplines. To improve the usefulness of existing scales, future studies are encouraged to adapt general feedback literacy scales to various disciplinary contexts, considering the unique characteristics of discipline-specific feedback practices, particularly in fields beyond the current focus on language writing.

6.3. Direct assessment and observable indicators

This review identified a striking pattern across the reviewed studies, most of which rely on single-source self-reported data to measure feedback literacy. While these scales (e.g., Song, 2022; Yu et al., 2022) provide access to learners' and teachers' perceptions, they risk conflating perceived abilities with actual feedback practices. For instance, Yu et al. (2022) acknowledged that their validation of the L2 Student Writing Feedback Literacy Scale relied solely on student self-reports, which may reflect perceived competencies rather than actual behaviours. They suggested future studies incorporate focus groups to validate the scale's alignment with observable feedback practices. Similarly, Song (2022) highlighted that those self-reported measures of student feedback literacy could be augmented by lecturer evaluations or analysis of feedback artefacts (e.g., revised drafts), offering a more objective perspective.

Future scale development could move beyond perceptions to include more behavioural indicators of feedback literacy or use direct assessment to measure how students react to situated feedback practices and think aloud about their underlying thoughts. Direct assessment evaluates behavioural data (e.g., recorded feedback interaction in feedback conferences or online discussions) and real-time engagement and affective responses (Carless & Winstone,2023b). By implementing this approach, researchers can mitigate biases and gain a more comprehensive understanding of feedback literacy as both a perceptual and practised construct.

6.4. Component design

After reviewing the structure and components of feedback literacy scales, the knowledge dimension needs to be considered for feedback literacy scale development, especially for SFL. While students' knowledge, including domain knowledge, task knowledge and feedback knowledge, is fundamental to their effective feedback engagement, it is often underrepresented in SFL. Only a few recent studies, such as Teng and Ma (2024) and Cui et al. (2025), have begun to incorporate dimensions like domain knowledge and feedback knowledge into their scales. Similarly, although three of the five TFL scales assessed teachers' knowledge, they placed less emphasis on teachers' domain knowledge, which is essential for the scales' responsiveness. Future scales could integrate the knowledge component to better capture how such knowledge underpins effective feedback engagement.

Students' capacities to elicit and provide feedback have been measured less on most existing scales. The traditional teacher-centred paradigm on feedback has focused on feedback reception and it often overlooks students' agentic roles in initiating feedback dialogues and providing feedback information. The growing attention on eliciting suggests a shifting recognition of students as active participants in feedback processes. For example, Zhan (2022) and Chen et al. (2024) explicitly include eliciting capacity, with the former naming it directly and the latter aligning it with communicative expression. Dawson et al. (2024) further highlight feedback-seeking as part of feedback behaviour. As feedback interactions increasingly occur in digital and AI-supported environments, where students can prompt tools (e.g., Generative AI) for feedback, the ability to elicit feedback becomes increasingly essential (Cui et al., 2025).

The capacity to provide feedback is even less represented. Although this skill plays a crucial role in contributing to reciprocal learning, it has been neglected, with only Dawson et al. (2024) explicitly addressing it as a standalone component. Providing feedback not only reinforces the providers' own understanding of quality standards but also promotes a culture of shared responsibility in learning communities (Molloy et al., 2020). Future SFL scales should ensure that students' eliciting and providing feedback capacities are fully recognised and measured.

It was found that the current TFL scales do not fully capture the teacher capabilities and dispositions needed to support the development of SFL. In the feedback process, both teachers and students share the responsibility for students' feedback uptake. The underlying premise in the feedback field is that TFL is intended to facilitate the development of SFL, as students' ability to benefit from feedback depends on teachers creating conducive environments for these outcomes (Nash & Winstone, 2017; Zhan, 2024). However, existing TFL scales lack certain capacities and dispositions necessary to foster the corresponding capacities and dispositions of SFL. Therefore, in the future TFL scale development, researchers need to contemplate which planning and follow-up strategies, and what commitment to follow up on students' revision, could facilitate students' uptake of feedback.

6.5. Psychometric measurement

The review of psychometric properties of feedback literacy scales highlights the need for using multiple methods to examine reliability and ensure generalizability, and use Rasch analysis as a complementary approach to factor analysis. Specifically, most scales chose to use only one method to examine the reliability of the scale by using Cronbach's alpha coefficient. However, Cronbach's alpha represents a lower bound to the reliability and, in many cases, underestimates the true reliability (Sijtsma, 2009). Therefore, it is recommended to use two or more methods to examine the reliability of the scales, such as composite reliability , McDonald's omega (ω), inter-item correlations, and item-total correlations. In addition, to ensure that test results could be applied to different populations, settings, and times. Multi-group CFA or Differential item functioning is recommended to to ensure the tool's effectiveness across various contexts. Lastly, Rasch analysis was less frequently used to examine the psychometric properties than factor analysis, which was widely used across the reviewed scales. Rasch analysis and factor analysis are complementary; using both together can enhance instrument refinement and offer a comprehensive interpretation of findings (Richardson, 2005; Yan, 2018). For instance, CFA provides numerous overall model fit indices, while Rasch analysis places more emphasis on the fit of each item to the model (Yan, 2016). Several empirical studies (e.g., ; Chang & Engelhard, 2016; Deneen et al., 2013; Hart et al., 2013; Primi et al., 2014) have demonstrated that both approaches yield unique insights and can jointly improve the validity and refinement of measurement instruments.

7. Conclusion

Overall, this review offers a thorough and detailed synthesis of the existing feedback literacy scales, highlighting their strengths and limitations. It could provide valuable insights to inform and guide the development of future measurement tools of feedback literacy in various educational settings. However, despite its comprehensive nature, the review has certain limitations that should be acknowledged.

Firstly, our literature search was confined to peer-reviewed journal articles, which means we did not include dissertations, theses, or other grey literature. This restriction might have led to the exclusion of

relevant and potentially innovative scales that are still unpublished or only available in non-peer-reviewed sources. As a result, some valuable tools or recent developments in feedback literacy measurement may have been overlooked. Secondly, there is a possibility that some psychometric evidence supporting the validity and reliability of the scales was underreported or not fully captured, largely because we relied on the authors' own reports. This dependence on existing publications means that if authors did not thoroughly report their psychometric data, some aspects of the scales' robustness might not be reflected in our synthesis. Finally, although we chose to follow the authors' original classifications of feedback literacy components for consistency and comparability, we recognise that these categorisations are open to interpretation and discussion.

Ethics approval

This review paper synthesises data from previously published studies. As the data analysed in this review are already publicly available, no additional ethics approval is required for this review.

Funding

This study was supported by General Research Fund (Reference No.:18605223) funded by University Grants Committee in Hong Kong.

CRediT authorship contribution statement

Ying Zhan: Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation, Conceptualization. **Zi Yan:** Writing – review & editing, Methodology, Conceptualization. **Ye Zeng:** Writing – original draft, Investigation, Data curation. **Wenyun Luo:** Writing – original draft, Investigation, Data curation. **Zhi Hong Wan:** Writing – review & editing, Writing – original draft, Supervision, Conceptualization.

Conflict of Interest

No potential conflict of interest was reported by the authors.

Data availability

Data will be made available on request.

References

The papers in our review are marked with an asterisk (*).

Biggs, J. (1996). Western Misperceptions of the Confucian-Heritage Learning Culture. In In. D. Watkins, & J. Biggs (Eds.), *The Chinese Learners: Cultural, Psychological and Contextual Influences* (pp. 45–68). The University of Hong Kong.

Carless, D., & Boud, D. (2018). The development of student feedback literacy: Enabling uptake of feedback. *Assessment Evaluation in Higher Education, 43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Carless, D., & Winstone, N. (2023a). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education, 28*(1), 150–163. https://doi.org/10.1080/13562517.2020.1782372

Carless, D., & Winstone, N. (2023b). *Feedback literacy: A practical guide to developing students' capabilities*. Routledge. https://doi.org/10.4324/9781003264627

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. https://doi.org/10.1177/0734282915593835

* Chen, H., Tang, S., Zhang, S., Xu, J., & Wang, G. (2024). Development and validation of the high school students' Mathematics Discourse Feedback Skills Scale (MDFSS). *Current Psychology, 43*, 30290–30305. https://doi.org/10.1007/s12144-024-06578-1.

* Cui, Y., Meng, Y., Qian, X., & Tang, L. (2025). Developing a college student feedback literacy scale for the GenAI context. *Assessment Evaluation in Higher Education*, 1–14. https://doi.org/10.1080/02602938.2025.2495048.

* Dawson, P., Yan, Z., Lipnevich, A., Tai, J., Boud, D., & Mahoney, P. (2024). Measuring what learners do in feedback: The feedback literacy behaviour scale. *Assessment Evaluation in Higher Education, 49*(3), 348–362. https://doi.org/10.1080/02602938.2023.2240983.

Deeley, S. J., Fischbacher-Smith, M., Karadzhov, D., & Koristashevskaya, E. (2019). Exploring the 'wicked'problem of student dissatisfaction with assessment and feedback in higher education. *Higher Education Pedagogies, 4*(1), 385–405. https://doi.org/10.1080/23752696.2019.1644659

Deneen, C., Brown, G. T. L., Bond, T. G., & Shroff, R. (2013). Understanding outcome-based education changes in teacher education: Evaluation of a new instrument with preliminary findings. *Asia-Pacific. Journal of Teacher Education, 41*, 441–456. https://doi.org/10.1080/1359866X.2013.787392

* Dong, Z., Gao, Y., & Schunn, C. D. (2023). Assessing students' peer feedback literacy in writing: Scale development and validation. *Assessment Evaluation in Higher Education, 48*(8), 1103–1118. https://doi.org/10.1080/02602938.2023.2175781.

Espasa, A., Mayordomo, R. M., Guasch, T., & Martinez-Melo, M. (2019). Does the type of feedback channel used in online learning environments matter? Students' perceptions and impact on learning. *Active Learning in Higher Education, 23*(1), 49–63. https://doi.org/10.1177/1469787419891307

Hajjar, S. T. (2018). Statistical analysis: Internal-consistency reliability and construct validity. *International Journal of Quantitative and Qualitative Research Methods, 6*(1), 27–38.

Han, Y., & Xu, Y. (2021). Student feedback literacy and engagement with feedback: A case study of Chinese undergraduate students. *Teaching in Higher Education, 26*(2), 181–196. https://doi.org/10.1080/13562517.2019.1648410

Hart, C. O., Mueller, C. E., Royal, K. D., & Jones, M. H. (2013). Achievement goal validation among African American high school students: CFA and Rasch results. *Journal of Psychoeducational Assessment, 31*(3), 284–299. https://doi.org/10.1177/0734282912466726

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

* Kang, C., Huang, J., Liu, Y., & Yin, H. (2025). Development and validation of a generic self-assessment scale for K-12 teachers as feedback givers: Insights from item response theory and factor analysis. *Humanities and Social Sciences Communications, 12*(1), 1–10. https://doi.org/10.1057/s41599-025-04927-4.

Lee, I. (2021). The development of feedback literacy for writing teachers. *TESOL Quarterly, 55*(3), 1048–1059. https://doi.org/10.1002/tesq.3012

* Lee, I., Karaca, M., & Inan, S. (2023). The development and validation of a scale on L2 writing teacher feedback literacy. *Assessing Writing, 57*, Article 100743. https://doi.org/10.1016/j.asw.2023.100743.

Little, T., Dawson, P., Boud, D., & Tai, J. (2024). Can students' feedback literacy be improved? A scoping review of interventions. *Assessment Evaluation in Higher Education, 49*(1), 39–52. https://doi.org/10.1080/02602938.2023.2177613⟩

Malecka, B., Boud, D., & Carless, D. (2022). Eliciting, processing and enacting feedback: Mechanisms for embedding student feedback literacy within the curriculum. *Teaching in Higher Education, 27*(7), 908–922. https://doi.org/10.1080/13562517.2020.1754784

Mandouit, L., & Hattie, J. (2023). Revisiting "The Power of Feedback" from the perspective of the learner. *Learning and Instruction, 84*, Article 101718. https://doi.org/10.1016/j.learninstruc.2022.101718

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American psychologist, 50*(9), 741. https://doi.org/10.1037/0003-066X.50.9.741

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine, 151*(4), 264. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Molloy, E., Boud, D., & Henderson, M. (2020). Developing a learning-centred framework for feedback literacy. *Assessment Evaluation in Higher Education, 45*(4), 527–540. https://doi.org/10.1080/02602938.2019.1667955

Nash, R. A., & Winstone, N. E. (2017). Responsibility-sharing in the giving and receiving of assessment feedback. *Frontiers in Psychology, 8*, 1–9. https://doi.org/10.3389/fpsyg.2017.01519

Ng, D. T. K., Wu, W., Leung, J. K. L., Chiu, T. K. F., & Chu, S. K. W. (2024). Design and validation of the AI literacy questionnaire: The affective, behavioural, cognitive and ethical approach. *British Journal of Educational Technology, 55*(3), 1082–1104. https://doi.org/10.1111/bjet.13411

Nieminen, J. H., & Carless, D. (2023). Feedback literacy: A critical review of an emerging concept. *Higher Education, 85*(6), 1381–1400. https://doi.org/10.1007/s10734-022-00895-9

Primi, R., Wechsler, S. M., Nakano, T. C., Oakland, T., & Guzzo, R. S. L. (2014). Using item response theory methods with the Brazilian Temperament Scale for students. *Journal of Psychoeducational Assessment, 32*(7), 651–662. https://doi.org/10.1177/0734282914528612

Richardson, J. T. (2005). Instruments for obtaining student feedback: A review of the literature. *Assessment Evaluation in Higher Education, 30*(4), 387–415. https://doi.org/10.1080/02602930500099193

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0

* Song, B. K. (2022). Bifactor modelling of the psychological constructs of learner feedback literacy: Conceptions of feedback, feedback trust and self-efficacy. *Assessment Evaluation in Higher Education, 47*(8), 1444–1457. https://doi.org/10.1080/02602938.2022.2042187.

Sutton, P. (2012). Conceptualizing feedback literacy: Knowing, being, and acting. *Innovations in Education and Teaching International, 49*(1), 31–40. https://doi.org/10.1080/14703297.2012.647781

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

* Teng, M. F., & Ma, M. (2024). Assessing metacognition-based student feedback literacy for academic writing. *Assessing Writing, 59*, Article 100811. https://doi.org/10.1016/j.asw.2024.100811.

* Wancham, K., & Tangdhanakanond, K. (2025). Development and validation of the student feedback literacy test for secondary schoolers. *Cogent Education, 12*(1). https://doi.org/10.1080/2331186x.2025.2481002.

* Wang, Y., Derakhshan, A., Pan, Z., & Ghiasvand, F. (2023). Chinese EFL teachers' writing assessment feedback literacy: A scale development and validation study. *Assessing Writing, 56*, Article 100726. https://doi.org/10.1016/j.asw.2023.100726.

Winstone, N. E., Balloo, K., & Carless, D. (2022). Discipline-specific feedback literacies: A framework for curriculum design. *Higher Education, 83*(1), 57–77. https://doi.org/10.1007/s10734-020-00632-0

* Woitt, S., Weidlich, J., Jivet, I., Göksün, D. O., Drachsler, H., & Kalz, M. (2023). Students' feedback literacy in higher education: An initial scale validation study. *Teaching in Higher Education*, 1–20. https://doi.org/10.1080/13562517.2023.2263838.

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

* Yang, Y., Luo, Z., Dong, Y., Kurup, P. M., & Wang, Y. (2023). Towards a new paradigm: The development and validation of a scale to explore technology-enhanced feedback literacy among primary and secondary school teachers. *Educational Technology Research and Development, 71*(2), 391–413. https://doi.org/10.1007/s11423-022-10168-y.

* Yildiz, H., e. Bozpolat, & e. Hazar. (2022). Feedback literacy scale: A study of validation and reliability. *International Journal of Eurasian Education and Culture, 7*(19), 2214–2249. https://doi.org/10.35826/ijoecc.624.

* Yu, S., Di Zhang, E., & Liu, C. (2022). Assessing L2 student writing feedback literacy: A scale development and validation study. *Assessing Writing, 53*, Article 100643. https://doi.org/10.1016/j.asw.2022.100643.

* Zhan, Y. (2022). Developing and validating a student feedback literacy scale. *Assessment Evaluation in Higher Education, 47*(7), 1087–1100. https://doi.org/10.1080/02602938.2021.2001430.

* Zhan, Y. (2024). Feedback Literacy of teacher candidates: Roles of assessment course learning experience and motivations for becoming a teacher. *The Asia-Pacific Education Researcher, 33*(5), 1117–1127. https://doi.org/10.1007/s40299-023-00779-1.

* Zhang, E. D., Zhou, N., & Yu, S. (2023). Assessing L2 secondary student writing feedback literacy and its predictive effect on their L2 writing performance. *Language Teaching Research*. https://doi.org/10.1177/13621688231217665.

Zhang, S., Xu, J., Chen, H., Jiang, L., & Yi, X. (2024). Influence of teacher autonomy support in feedback on high school students' feedback literacy: The multiple mediating effects of basic psychological needs and intrinsic motivation. *Frontiers in Psychology, 15*, Article 1411082. https://doi.org/10.3389/fpsyg.2024.1411082