Routledge
Taylor & Francis Group

OPEN ACCESS

Check for updates

The student feedback literacy instrument (SFLI): multilingual validation and introduction of a short-form version

J. Weidlich[a,b,c] (ID), I. Jivet[c,d] (ID), S. Woitt[e], D. Orhan Göksün[f] (ID), J. Kraus[g] (ID) and H. Drachsler[c,h,i] (ID)

[a]University of Zürich, Zürich, Switzerland; [b]Zürich University of Teacher Education, Zürich, Switzerland; [c]DIPF – Leibniz-Institute for Research and Information in Education, Frankfurt, Germany; [d]FernUniversität in Hagen, Hagen, Germany; [e]Heidelberg University of Education, Heidelberg, Germany; [f]Adyiaman University, Adıyaman, Turkey; [g]Johannes Gutenberg University of Main, Mainz, Germany; [h]Goethe University, Frankfurt, Germany; [i]Open University of the Netherlands, Heerlen, The Netherlands

ABSTRACT

Feedback literacy is gaining recognition as a key concept for understanding how engage with and learn from feedback in higher education. This study presents validity evidence for a refined version of the Student Feedback Literacy Instrument (SFLI), designed to measure the construct across two dimensions—feedback attitudes and feedback practices—in German, English, and Turkish. We developed both a full-length and a short-form version (SFLI-S). Using confirmatory factor analyses on different student samples (N_{total}= 1424), we confirmed the two-factor structure across languages, supporting the model of feedback literacy comprising of attitudinal and behavioral components. Associations with related constructs further support the instrument's convergent validity. As a psychometrically sound, multilingual instrument, the SFLI can facilitate cross-cultural feedback literacy research and provide a valuable tool for research and educational practice. The SFLI-S offers an economical alternative, enabling wider integration into studies on how students engage with feedback.

Introduction

Feedback is widely recognized as a crucial factor in learning (Hattie and Timperley 2007). Its importance is underscored by its central role in models of self-regulated learning (Butler and Winne 1995) and its potential to significantly impact student outcomes. However, the effectiveness of feedback can vary considerably (Wisniewski, Zierer, and Hattie 2019) and recent research emphasizes a student-centered approach, positioning *student feedback literacy* as a crucial construct to make the most of feedback in higher education.

Student feedback literacy, initially introduced by Sutton (2012) and further developed by Carless and Boud (2018), refers to the "understandings, capacities and dispositions needed to make sense of information and use it to enhance work or learning strategies" (p. 1316). This concept emphasizes students' agency in utilizing feedback to improve their learning strategies

and achieve their educational goals (see e.g. Molloy et al., 2020). In our previous work (Woitt et al. 2025), we responded to calls for a psychometrically sound measurement of feedback literacy (e.g. Winstone, Mathlin, and Nash 2019) by developing and validating a preliminary self-report instrument. This initial instrument provided researchers with a tool to capture this construct empirically, facilitating more in-depth investigations into feedback literacy in higher education.

This study builds on our previous work by refining and validating the Student Feedback Literacy Instrument (SFLI). First, we propose modifications to the original instrument, evaluating these against the initial version. Second, we develop a short form (SFLI-S) with sound psychometric properties and improved survey economy. Third, we provide validated translations of the SFLI and SFLI-S in German, English, and Turkish to support cross-cultural and -lingual research (Rovagnati and Pitt 2022; Pazio Rossiter and Bale 2023). Lastly, we assess the psychometric properties of all versions and provide convergent validity evidence by examining their associations with established educational constructs.

Through this instrument development effort, we contribute to theoretical scholarship by empirically confirming a conceptually intuitive yet unique factor structure—comprising only two main dimensions, attitudinal dispositions and behavioral tendencies—that stands out within the existing feedback literacy literature. Further, a central contribution is that we equip researchers and educators with a psychometrically sound and cross-culturally applicable instrument tool for measuring feedback literacy in students. We are optimistic that the short-form instrument will enable even more widespread integration of the feedback literacy instrument to measure and analyze student feedback literacy in diverse research contexts and educational practice.

Available measures

Several psychometric instruments have been developed to measure student feedback literacy. In developing our initial instrument (originally published in May 2023, but now included in the 2025 volume of the journal, Woitt et al. 2025), we reviewed three general feedback literacy scales (Song 2022; Yildiz, Bozpolat, and Hazar 2022; Zhan 2022), whereas Dawson et al. (2024) was published while our work was in press. Specific adaptations, like those for second-language writing (Yu, Di Zhang, and Liu 2022) or peer feedback literacy (Dong, Gao, and Schunn 2023), are excluded as they target specific domains, whereas our instrument is intended to be more general-purpose.

We identified two main limitations in existing instruments: omission of behavioral aspects and premature confirmatory approaches (Woitt et al. 2025). First, we observed that although these instruments (e.g. Song 2022; Yildiz, Bozpolat, and Hazar 2022) capture a breadth of attitudes and dispositions, some omit behavioral aspects; for example, whether students tend to enact in subsequent learning what they have gleaned from feedback. We (Woitt et al. 2025) and Dawson et al. (2024) thus argued that a suitable feedback literacy scale should incorporate students' behaviors, that is, students' habits, tendencies, practices as they relate to feedback. Theoretical frameworks, such as Carless and Boud (2018) dimensions, underscore the importance of including behavioral aspects like *taking action*.

A second concern is methodological. Except for Yildiz, Bozpolat, and Hazar (2022), most scale development efforts relied exclusively on confirmatory factor analysis (CFA), which assumes robust theoretical grounding. And indeed, Dawson et al. (2024) and Song (2022) justify their confirmatory approach by referring to key theoretical and conceptual papers. However, given the relatively young literature on feedback literacy, we deemed this assumption premature, evidenced by the fact that the scale developers gleaned noticeably diverging a priori factor structures from the same literature. Arguably, this betrays that conceptual foundations of feedback literacy may not be as unambiguous as hoped. Considering this, we opted for a more open-ended approach in our scale development to allow for one of many potential empirical models arising from our

initial assumptions (Woitt et al. 2025). The following section summarizes our approach, what was found, and the describes the need for ongoing validation research.

Initial student feedback literacy instrument (Woitt et al. 2025)

Development and Results

We created a comprehensive item pool, initially covering 83 descriptions of feedback literacy in the literature, consolidated into 16 subthemes, and later refined to 11 key facets (Woitt et al. 2025, OSF: osf.io/z4tus). These facets allowed us to operationalize feedback literacy by generating items representing each facet. We suggested that the facets could be grouped into three overarching clusters, i.e. dimensions: *openness to feedback*, *engagement with feedback*, and *enactment of feedback*. Openness captures an understanding of feedback and productive attitudes toward feedback. Engagement refers to the depth of students' feedback processing and their feedback-seeking behaviors. Finally, enactment concerns the application of feedback for learning improvements or performance understanding (see Supplemental Material A).

Crucially, despite expecting a three-factor structure based on the derived facets, in Woitt et al. (2025), we used exploratory factor analysis (EFA) to allow alternative factor structures. This approach was complemented by principled analysis steps, e.g. a priori decision rules for item removal (0.4-0.3-0.2 rule and 0.32 threshold, Howard 2016; Costello and Osborne 2005). Although a three-factor structure was anticipated, exploratory factor analysis (EFA) revealed a two-factor model. Engagement and enactment merged into a single behavioral factor, while the appraisal facet shifted from engagement to openness. The two dimensions were labeled *feedback attitudes* (e.g. "I feel responsible for using feedback") and *feedback practices* (e.g. "I refine my learning strategies based on feedback"). Rasch analyses confirmed the instrument's psychometric soundness the 21 items initial instrument (see table 1 in Woitt et al. 2025 for the list of items representing facets and dimensions).

Need for further research

While this instrument is usable for research—and indeed has been used already (see Weidlich et al. 2025)—more validation work is needed. Exploratory research should ideally be followed by a confirmatory approach to ensure that the findings generalize. Further, from a conceptual perspective, the two-factor structure diverges notably from other scale development efforts by positing a more parsimonious model (i.e. only two dimensions). While our model still fundamentally aligns with seminal conceptions (e.g. Carless and Boud 2018) by encompassing students' beliefs and attitudes and behavioral components, confirmatory evidence is required to claim validity of the instrument.

Further development is needed to address imbalances in the initial instrument. Iterative item removal resulted in some facets being represented by a single item (e.g. *decoding*), while others had up to three items (e.g. *agency*). Since there is no theoretical justification for this asymmetry, it is important to broaden the underrepresented facets while keeping the instrument concise by avoiding excessive representation of other facets. Accordingly, we propose a revised version of the Woitt et al. (2025) instrument with more balanced facet representation.

This research also seeks to enhance the instrument's utility across diverse contexts. First, the 21-item version may be too lengthy for use in settings with multiple measurement points or extensive surveys, risking survey fatigue and poor response quality (Porter, Whitcomb, and Weitzer 2004; Rolstad, Adler, and Rydén 2011). Other available instruments range from 21 to 24 items (Song 2022; Yildiz, Bozpolat, and Hazar 2022; Zhan 2022; Dawson et al. 2024), highlighting the need for a more concise instrument with solid psychometric properties. Second, we aim to

improve accessibility by translating the scale into additional languages and evaluating its psychometric properties. To summarize, we derive the following research aims for this study:

1. Evaluate a revised instrument and confirm the psychometric properties of the SFLI.
2. Identify and confirm a short version of the instrument.
3. Establish the psychometric properties of translations of the instrument.

Method

Supplemental material for this research can be found at the Open Science Framework (OSF): https://osf.io/e73jy/.

Translation

For the English-German translation, two translators independently translated the items: one with native-level proficiency in both languages and another a native German speaker with strong English skills. A third translation was generated using DeepL, a neural machine translation tool specialized in academic and European languages (MachineTranslation.com 2023). The three versions were reviewed by two co-authors to select the most accurate and fluent phrasing. DeepL provided the best translation for one-third of the items. The Turkish translation followed the same process as described in Woitt et al. (2025), involving forward-back translations by a language expert unfamiliar with the original items.

Samples

We obtained data from four samples in three languages: German, English, and Turkish. Final sample sizes ranged from $N = 225$ (English sample) to 453 (First German sample), and they consisted of a majority of women (53–81%) and few non-binary participants. An overview of the samples is provided in Table 1. More detailed information about the samples and data cleaning procedures are provided in Supplemental Material B.

Analytic approach

While our goal was ultimately confirmatory, we did not want to preclude the option of modifying the instrument according to the data. This approach is sometimes called EFA-in-CFA framework (E/CFA, Brown 2015) and offers some flexibility for ad hoc modifications such as improving model fit or removing problematic items. To prevent overfitting and ensure model generalizability, we provide further confirmatory evidence *via* CFA in an independent sample as much as possible. For the German SFLI, we were able to collect such additional independent data (see section

Table 1. Sample overview.

	German #1	German #2	English	Turkish
Source	Online panel Bilendi GmbH	Higher education lecture class	Online sampling prolific. com	Higher education course
N Before cleaning	526	530	283	297
Final sample N	453	512	225	234
Gender	56.3% women (three non-binary persons)	81% women (two non-binary persons)	53% women (four non-binary persons)	70% women (no non-binary persons)
Mean Age (SD)	25.28 (5.73) years	Not available	23.9 (4.3) years	22.4 (3.8) years

"Samples") and thus can provide strictly confirmatory evidence. For the English and Turkish SFLI, however, only one sample each was available, making these instrument versions more preliminary and requiring future confirmatory studies. These limitations are noted in the results section to clarify the strength of evidence.

Model fit was assessed with multiple fit indices following established benchmarks presented by Hu and Bentler (1999) and Schreiber (2008). In addition to the chi-square statistic, which is too sensitive to large samples, we report the more flexible X^2/df ratio (adequate fit ≤ 3; good fit ≤ 2), root mean square error of approximation (RMSEA) and standardized root mean square residual (SRMR) (adequate fit ≤ 0.08; good fit ≤ 0.05), and Tucker-Lewis index (TLI) and comparative fit index (CFI) (adequate fit ≥ 0.90; good fit ≥ 0.95). Direct model comparisons relied on Akaike information criterium (AIC) and Bayesian information criterium (BIC), where lower values indicate a better fit.

For all factor models aside from the SFLI-S, residual covariances between items within a facet were specified a priori. In a default CFA model, it is assumed that covariance among items arises entirely from the latent factor, and error terms are uncorrelated. Based on our understanding of the underlying theory, however, we can predict residual correlations for items from the same facet. See section "Refinement and confirmation of SFLI (Goal 1)" for more details.

To ensure that the instruments are unbiased, we assessed differential item functioning (DIF) in each sample. An instrument is unbiased when it works the same way for different people, for example, when it measures the same thing in men and women, or younger and older students. The available student characteristics (e.g. gender, age) varied across the convenience samples used in different research projects. For example, migration status was assessed in one German sample, while the field of study was recorded in another. Thus, we were able to assess DIF through a variety of potentially relevant parameters.

For additional validity evidence, we evaluate the convergent validity of SFLI with distinct but plausibly related instruments. The constructs were collected in research projects not directly related to the instrument development goals. As such, the available constructs differ between samples. For instance, the first German sample included data on students' need for cognition (NFC; Cacioppo et al. 1996), while the second included grit (Duckworth and Quinn 2009). Established constructs served as references for evaluating convergent validity.

Results

### *Refinement and confirmation of SFLI (Goal 1)*

The first model, Model A, represents the two-factor instrument developed by us in Woitt et al. (2025), which comprises 21 items across two dimensions: *feedback attitudes* (9 items) and *feedback practices* (12 items). This model is displayed in Figure 1 (top). As outlined in the section "Initial Student Feedback Literacy Instrument (Woitt et al. 2025)", the factor structure was derived through an exploratory approach, where items were removed based on inconsistent loadings or unsuitability for EFA. This process resulted in certain facets of the construct being underrepresented (e.g. only one item for *appraisal*), while others were comparatively overrepresented (e.g. three items for *agency*).

To address these imbalances and improve content validity, we developed Model B, which aimed for equal representation across facets, with two items per facet. This approach balanced survey economy and construct breadth. Additional items for underrepresented facets were developed following the same procedure outlined in Woitt et al. (2025), i.e. theoretical alignment and expert review. For overrepresented facets items with least unique content to its respective facet were removed. Thus, Model B consists of 22 items (8 for feedback attitudes, 14 for feedback practices; Figure 1, middle).
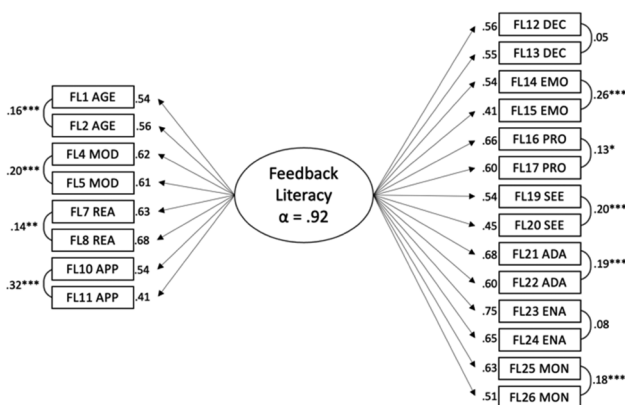
For comparison purposes, we also specified a single-factor feedback literacy model (Model C, Figure 1, bottom) to evaluate whether the complexity introduced by multiple factors was

Model A (Woitt et al., 2025)   Fit indices
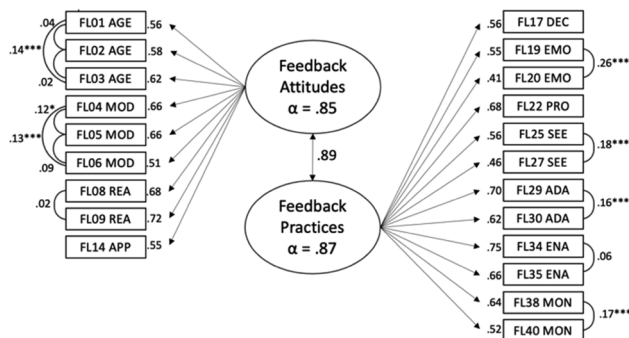
Model B (revised SFLI)   Fit indices
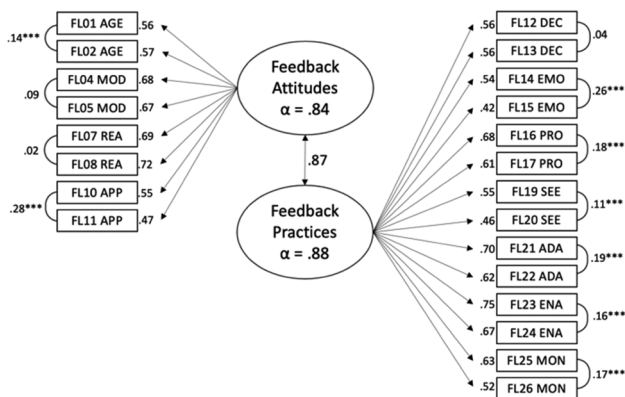
Model C (single factor model)   Fit indices

Figure 1. Psychometric properties of Model A (top), B (middle), and C (bottom). Note that indicator names in Model A follow the item numbering of Woitt et al. (2025), whereas Model B follows an updated numbering. Letters accompanying item numbers refer to their respective facet.

justified. This simplest model allows us to benchmark whether a two-dimensional factor structure is needed in the first place or if a less nuanced but even more parsimonious alternative explains the data better.

Model specifications for Models A and B included correlated residuals to account for shared variance within facets, as suggested by the theoretical facet structure. For example, in Model A, the three agency items had correlated errors, while Model B specified one correlated residual path for each two-item facet (Figure 1). This ensured the models accurately reflected the theoretical structure. The analyses were conducted on the first German-language sample (see section "Samples").

Analyses revealed that, for all three models, items loaded significantly on their respective factors ($p < 0.001$) with loadings > 0.40 (see Figure 1). Model A and Model B exhibited strong fit indices, but Model B demonstrated superior fit across relative Chi-Square (X^2/df), CFI, TLI, SRMR, and RMSEA despite higher AIC and BIC values. Both models fit the data better than the single-factor Model C, which showed poorer global fit and more severe residual covariances. Supplemental material C includes residual and modification indices as additional fit statistics.

Model B exhibited fewer unsystematic residual covariances and areas requiring modification compared to Model A. While Model A fit well, the more balanced and theoretically aligned Model B offers superior psychometric properties. Consequently, we reject Model C and favor Model B over Model A, given its better overall fit and substantive balance. Internal consistency reliability for Model B was α = 0.84 (feedback attitudes) and α = 0.88 (feedback practices), with the factors correlating strongly, $r = .87$. To test robustness, all models were rerun without correlated errors, which reduced fit indices universally but did not affect Model B's relative superiority (see Supplemental Material D).

We validated Model B using a fully confirmatory analysis in the second German sample, which yielded strong fit indices: $X^2(197) = 416.48$, $p < 0.001$; $X^2/df = 2.11$; CFI = 0.93; TLI = 0.92; SRMR = 0.05; RMSEA = 0.05 (90% CI: 0.04–0.05); AIC = 24,579.5; BIC = 24,910.09. Item loadings were significant at $p < 0.001$, except for FL10 ("When evaluating feedback, I take into account that it was shaped by the provider's perspective"), which loaded below .40. Internal consistency was α = 0.74 for feedback attitudes and α = 0.85 for feedback practices, with a factor correlation of $r = .78$. Based on this evidence, we designate this model as the Student Feedback Literacy Instrument (SFLI). The final set of items for the SFLI is listed in Table 2.

We assessed DIF for the new SFLI using demographic indicators across the German samples. Analyses were conducted separately for the dimensions of feedback attitudes and feedback practices. Included indicators were gender and socioeconomic status (SES). Gender was binary-coded (man, woman) due to insufficient data for non-binary categories. Conditional Likelihood Ratio (CLR) tests showed no significant DIF for feedback attitudes, CLR (42) = 55.40, $p = 0.081$, but significant DIF for feedback practices, CLR (77) = 187, $p < .001$. Partial gamma coefficients indicated that men scored higher on FL15 ("I handle feedback on a factual level instead of taking it personally," PGC = 0.36, $p < 0.001$). Using the McArthur Scale (Hoebel et al. 2015), SES was categorized into high (>7; $n = 196$), middle (≤7; $n = 132$), and low (<5; $n = 181$). CLR tests revealed no significant DIF for feedback attitudes, CLR (62) = 57.30, $p = 0.64$, or feedback practices, CLR (110) = 130, $p = .093$. These findings suggest minor DIF in feedback practices related to gender but no systematic bias across SES categories.

Development of a short form of the instrument (Goal 2)

To create a short version of the SFLI (SFLI-S), we selected one item per facet. Selection criteria included factor loadings, local model strain, and substantive reasoning (i.e. whether the item adequately represented the facet). When criteria conflicted, final decisions were made based on substantive considerations. The feedback attitudes dimension, consisting of four facets, was represented by four items: FL2 (agency, λ = 0.57), FL4 (model, λ = 0.68), FL8 (readiness, λ = 0.72), and FL10 (appraisal, λ = 0.55). The feedback practices dimension, comprising seven facets, was represented by seven items: FL13 (decoding, λ = 0.56), FL14 (emotion, λ = 0.54), FL16 (processing, λ = 0.68), FL19 (seeking, λ = 0.55), FL21 (adaptation, λ = 0.70), FL23 (enactment, λ = 0.75), and

Table 2. Summary of the assessed instruments.

	German		English		Turkish	
	SFLI	SFLI-S	SFLI	SFLI-S	SFLI	SFLI-S
Feedback attitudes (α)	.74	.68	.73	.65	.77	.64
FL1 (agency): I think that a feedback process is most effective if I take an active role in it.	X	X	X	X	X	X
FL2 (agency): I believe that I can contribute to the value of feedback processes.	X	X	X	X	X	X
FL4 (model): I believe that one of the main purposes of feedback is for me to improve in my studies.	X		X		X	
FL5 (model): I am convinced that working through feedback makes me better at evaluating my own work.	X		X		X	
FL7 (readiness): I am interested in receiving feedback about my learning.	X	X	X		X	X
FL8 (readiness): I am committed to making the most of feedback to succeed in my studies.	X	X	X		X	
FL10 (appraisal): When evaluating feedback, I take into account that it was shaped by the providers' perspective.	X	X	X	X	X	X
FL11 (appraisal): I believe that people with different perspectives will give me different feedback.	X		X	X	X	
Feedback practices (α)	.85	.72	.85	.77	.83	.72
FL12 (decoding): If needed, I seek out further information to better understand a feedback comment.	X		X			
FL13 (decoding): I always manage to get valuable information out of the feedback I receive.	X	X	X	X	X	X
FL14 (emotion): When dealing with feedback, I try to keep my emotional balance.	X		X	X	X	X
FL15 (emotion): I handle feedback on a factual level instead of taking it personally.	X		X		X	
FL16 (processing): I take all the time I need to reflect on feedback I have received.	X	X	X	X	X	
FL17 (processing): When I receive feedback, I carefully take note of every comment.	X		X		X	
FL19 (seeking): I assess my learning progress to determine where feedback might be helpful to me.	X	X	X	X	X	X
FL20 (seeking): I always consult multiple sources of feedback to obtain diverse perspectives.	X		X		X	
FL21 (adaptation): I conclude from feedback how to do things in the future.	X	X	X	X	X	X
FL22 (adaptation): I reconsider and refine my learning strategies based on feedback.	X		X		X	
FL23 (enactment): I strive to make the most of the feedback I have received.	X	X		X	X	X
FL24 (enactment): If given the opportunity, I always revise my work based on feedback.	X		X		X	
FL25 (monitoring): I consistently use feedback as a reference point to judge my overall progress.	X	X	X	X	X	X
FL26 (monitoring): I take feedback into account for evaluating how well I am navigating a challenge.	X		X		X	
Factor correlation r	.78	.89	.60	.76	.68	.75
CFI	.93	.94	.93	.94	.92	.95
TLI	.92	.92	.92	.92	.90	.93
SRMR	.04	.04	.05	.05	.06	.05
RMSEA	.05	.05	.05	.06	.05	.05

Note. X denotes the item's inclusion in the final factor model, whereas the grey background shows the established factor structure of the SFLI *via* Model B. X on a white background or, vice versa, a grey background sans X denotes ad-hoc modifications.

FL25 (monitoring, λ = 0.63). This model demonstrated good fit: $X^2(43)$ = 100.31, $p < 0.001$; $X^2/df = 2.33$; CFI = 0.96; TLI = 0.95; SRMR = 0.03; RMSEA = 0.05 (90% CI: 0.04–0.07). All items loaded significantly on their respective factors, and the factor correlation was $r = .93$. Internal consistency reliability was α = 0.71 for feedback attitudes and α = 0.82 for feedback practices.

To confirm the psychometric properties of the SFLI-S, we tested the model in the second German sample with identical specifications. This analysis yielded a well-fitting model: $X^2(43)$ = 98.31, $p < 0.001$; $X^2/df = 2.29$; CFI = 0.94; TLI = 0.92; SRMR = 0.04; RMSEA = 0.05 (90% CI: 0.04–0.06). While two items (FL10 and FL19) did not meet the 0.40 loading threshold (λ = 0.26 and λ = 0.36, respectively), all items loaded significantly. The factor correlation was $r = 0.89$, and internal consistency reliability was α = 0.68 for feedback attitudes and α = 0.72 for feedback practices.

For a detailed overview of the SFLI and SFLI-S in all three languages, see Table 2.

English-language instrument (Goal 3a)

To evaluate the English translation of the SFLI, we conducted a CFA using the same specifications as Model B, including correlated errors between items within a facet. The model fit was reasonable but did not meet all thresholds: $X^2(197)$ = 337.56, $p < 0.001$; $X^2/df = 1.71$; CFI = 0.89; TLI = 0.87; SRMR = 0.07; RMSEA = 0.06 (90% CI: 0.05–0.07). Factor loadings ranged from 0.30 (FL10) to 0.71 (FL7), with FL10 and FL23 flagged as problematic based on modification indices (33.65 and 17.7, respectively). Removing FL10 improved fit significantly: $X^2(178)$ = 278.91, $p < 0.001$; $X^2/df = 1.56$; CFI = 0.92; TLI = 0.90; SRMR = 0.05; RMSEA = 0.05 (90% CI: 0.04–0.07). Removing FL23 further enhanced model fit: $X^2(160)$ = 238.56, $p < 0.001$; $X^2/df = 1.49$; CFI = 0.93; TLI = 0.92; SRMR = 0.05; RMSEA = 0.05 (90% CI: 0.03–0.06). These modifications left two facets represented by only one item each: FL11 (appraisal) and FL24 (enactment). Despite this, the instrument retains sufficient construct breadth and can serve as a preliminary model. Factor correlations were $r = 0.60$, and internal consistency reliability was α = 0.73 (feedback attitudes) and α = 0.85 (feedback practices). Given these exploratory modifications, these analyses should be considered partly exploratory instead of strictly confirmatory, making the findings preliminary and pending future confirmation.

DIF was assessed using gender and education as demographic indicators. Gender was binary-coded due to insufficient non-binary cases. CLR tests indicated no significant DIF for feedback attitudes, CLR(27) = 39.60, $p = 0.055$, or feedback practices, CLR (50) = 59.4, $p = .170$. Similarly, no DIF was observed across educational levels (undergraduate, graduate, doctorate) for feedback attitudes, CLR (54) = 43.30, $p = 0.851$, or feedback practices, CLR (100) = 115.60, $p = 0.140$.

CFA on the English SFLI-S suggested poor model fit: $X^2(43)$ = 101.41, $p < 0.001$; $X^2/df = 2.36$; CFI = 0.88; TLI = 0.85; SRMR = 0.06; RMSEA = 0.08 (90% CI: 0.06–0.10). FL10 showed high modification indices (24.38), and residual covariances between FL8, FL23, and FL10 indicated local misfit. However, no substantive basis exists for specifying correlated errors across facets, leaving the misfit unresolved. As FL10 is the only item representing the appraisal facet, its removal would compromise construct validity. To address this, we replaced FL10 with FL11, preserving the appraisal facet. This modification improved model fit: $X^2(43)$ = 73.52, $p < 0.003$; $X^2/df = 1.7$; CFI = 0.94; TLI = 0.92; SRMR = 0.05; RMSEA = 0.06 (90% CI: 0.03–0.08). This adjustment produced a usable preliminary SFLI-S in English, pending rigorous confirmatory testing. Internal consistency reliability was α = 0.65 (feedback attitudes) and α = 0.77 (feedback practices), with factor correlation $r = 0.75$.

Turkish-language instrument (Goal 3b)

A CFA of the Turkish translation of the SFLI, based on Model B specifications, revealed insufficient model fit: $X^2(197)$ = 382.4, $p < 0.001$; $X^2/df = 1.94$; CFI = 0.87; TLI = 0.85; SRMR = 0.06;

RMSEA = 0.06 (90% CI: 0.05–0.07). To improve the model, problematic items were removed iteratively while ensuring that each facet retained at least one representative item. Item FL16 was removed due to a non-significant factor loading, followed by FL12, which showed a high modification index (MI = 22.18) related to its non-significant loading and residual covariance with FL11. Finally, FL8 was removed due to its high modification index (MI = 9.51) and residual covariance with FL23. After these adjustments, the model fit improved substantially: $X^2(143)$ = 236.58, $p < 0.001$; $X^2/df = 1.66$; CFI = 0.92; TLI = 0.90; SRMR = 0.06; RMSEA = 0.05 (90% CI: 0.04–0.06). However, the modifications resulted in the facets readiness, decoding, and processing being represented by only one item each, reducing the depth of coverage for these dimensions. Consequently, this analysis should be considered partly exploratory (E/CFA), and further refinement of the instrument is needed to ensure that all facets of feedback literacy are adequately represented. A strictly confirmatory evaluation of the psychometric properties of the Turkish SFLI should follow future revisions. Internal consistency reliability for the instrument was α = 0.77 for feedback attitudes and α = 0.83 for feedback practices, with a factor correlation of r = .68. While the modified instrument can serve as a preliminary version, caution is advised when using it in research contexts.

DIF was assessed using gender and field of study as demographic indicators. Gender was binary-coded due to insufficient non-binary cases. CLR tests for gender indicated no significant DIF for feedback attitudes, CLR (19) = 28.4, $p = 0.076$, or feedback practices, CLR (25) = 31.1 p = .18. Similarly, the number of semesters enrolled did not lead to DIF for feedback attitudes, CLR (57) = 67.77, $p = 0.16$, nor feedback practices, CLR (75) = 80.80, $p = 0.30$.

For the Turkish translation of the SFLI-S, the initial CFA indicated reasonable but non-ideal fit: $X^2(43)$ = 80.79, $p < 0.001$; $X^2/df = 1.87$; CFI = 0.91; TLI = 0.88; SRMR = 0.05; RMSEA = 0.06 (90% CI: 0.03–0.08). Consistent with the findings for the full instrument, FL16 showed a non-significant factor loading, presenting a challenge for retaining the processing facet. To address this, FL16 was replaced with FL17, an item that loaded adequately in the long-form instrument. Similarly, FL8 exhibited large residual covariances with FL10 and FL23, making it a candidate for removal. To maintain representation of the readiness facet, FL8 was replaced with FL7, another item from the long-form SFLI. These modifications resulted in a model with good fit: $X^2(43)$ = 64.79, $p < 0.001$; $X^2/df = 1.51$; CFI = 0.95; TLI = 0.93; SRMR = 0.05; RMSEA = 0.05 (90% CI: 0.02–0.07). While these changes produce a preliminary SFLI-S with adequate psychometric properties in Turkish, the modifications require rigorous confirmatory testing in future research. Internal consistency reliability for the SFLI-S was α = 0.64 for feedback attitudes and α = 0.72 for feedback practices, with a factor correlation of $r = 0.75$.

Convergent validity of SFLI instruments

In the context of establishing psychometric properties of the instruments, we also assessed convergent validity, that is, the extent to which the instrument is associated with plausibly related constructs measured by well-established instruments. In the following, we assess convergent validity with the constructs need for cognition, grit, feedback self-efficacy, and surface and deep motivation.

Need for cognition predicting feedback literacy

Need for cognition (NFC) reflects an individual's preference for engaging in complex, effortful activities (Cacioppo et al. 1996). This trait is particularly relevant in educational contexts, as it influences how students process information and solve problems. Empirical evidence links higher NFC to greater engagement with complex tasks and improved academic achievement (Liu and Nesbit 2024). Given that feedback literacy involves effortful processing and active engagement with feedback, we hypothesized that NFC would positively predict feedback literacy. Specifically,

students with higher NFC were expected to value detailed feedback and use it to improve their learning outcomes.

NFC was measured in the first German sample using an 8-item shortened version of the German NFC scale by Bless et al. (1994), adapted by Kraus et al. (2019) for representativeness based on difficulty and factor loadings. Example items include: "I enjoy finding new solutions to problems," with responses rated on a 7-point Likert scale from (1) "does not apply at all" to (7) "applies completely."

To test our hypothesis, we fit a structural equation model (SEM) with robust maximum likelihood estimation. Feedback literacy was modeled as the dependent (endogenous) factor and NFC as the independent (exogenous) factor, based on the premise that NFC, as a general trait, influences the more context-specific processes underlying feedback literacy. Residual covariances for item pairs within feedback literacy facets accounted for shared variance. The model demonstrated reasonable fit: $X^2(391) = 837.38$, $p < 0.001$; $X^2/df = 2.14$; CFI = 0.90; TLI = 0.89; SRMR = 0.06; RMSEA = 0.05 (90% CI: 0.05, 0.05). Most factor loadings were significant ($p < 0.001$) and exceeded 0.40, though one NFC item showed a lower standardized estimate ($\lambda = 0.24$). Path coefficients from NFC to both feedback literacy dimensions were positive, significant, and of moderate effect size (Cohen 1988), with overlapping confidence intervals indicating similar predictive strength for both dimensions (see Figure 2).

Grit predicting feedback literacy

Grit, defined as perseverance and passion for long-term goals (Duckworth et al. 2007), is linked to academic achievement beyond cognitive ability, particularly in challenging contexts. Research indicates that individuals with high grit sustain effort and interest over time (Hagger and Hamilton 2019; Lam and Zhou 2022). We hypothesized that grit would predict both dimensions of feedback literacy, as individuals with grit are more likely to engage persistently with feedback to improve performance. The feedback literacy items generally reflect a purposeful attitude and effortful behavior toward feedback, representing feedback-specific expressions of grit's underlying mindset (feedback attitudes) and behaviors (feedback practices).

This analysis used data from the second German sample with an 8-item Grit scale (Schmidt et al. 2019) consisting of "sustained interest" (negatively worded, e.g. "I have trouble staying focused on projects that take months") and "perseverance" (e.g. "Whenever I start something new, I also complete it") dimensions. We expected negative correlations between feedback literacy
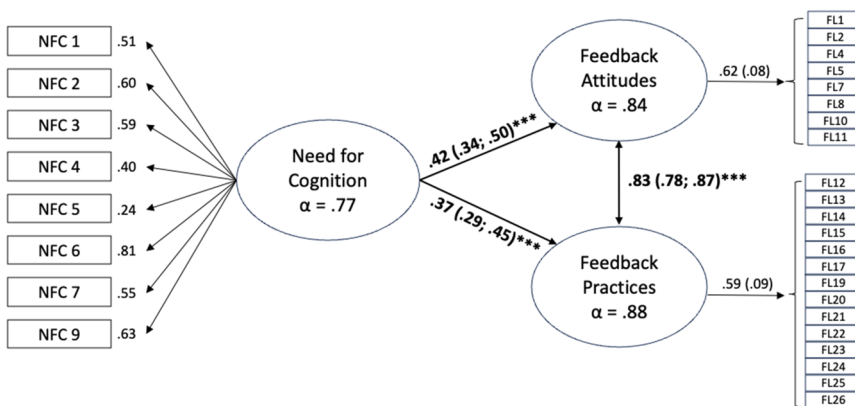

Figure 2. SEM model estimating paths of the need for cognition construct to feedback literacy dimensions. Mean factor loadings and standard deviation of feedback literacy indicators are shown to reduce visual clutter. Parentheses of path coefficients report 95% confidence intervals. *** $p < 0.001$.

Figure 3. SEM estimating paths of grit dimensions to feedback literacy dimensions. Mean factor loadings and standard deviation of feedback literacy indicators are shown to reduce visual clutter. Parentheses of path coefficients report 95% confidence intervals. ** $p < 0.01$; *** $p < 0.001$.
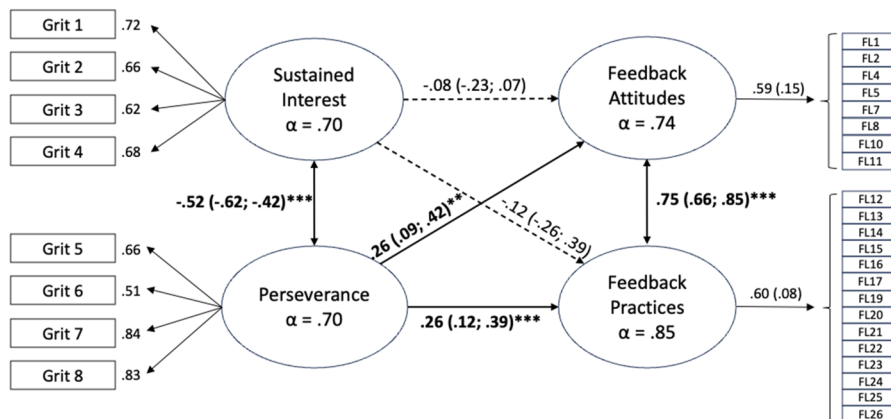
dimensions and sustained interest and positive correlations with perseverance. Grit was measured on a 5-point Likert scale from (1) "does not apply at all" to (5) "applies completely."

We fit a SEM with feedback literacy as the dependent variable and grit dimensions as independent variables, assuming that grit's broader trait scope relative to feedback literacy makes this causal direction more plausible than vice versa. For feedback literacy factors we specified residual covariances between item pairs of a facet. The model fit reasonably well: $X^2(388) = 791.97$, $p < 0.001$; $X^2/df = 2.04$; CFI = 0.90; TLI = 0.89; SRMR = 0.05; RMSEA = 0.05 (90% CI: 0.04, 0.05). Factor loadings were all significant ($p < 0.001$) and above 0.4, except for FL10 ($\lambda = 0.24$). Contrary to expectations, only two of the four paths were significant: perseverance showed small to moderate positive paths to feedback attitudes and practices, while paths from sustained interest were negative but not statistically significant. Additionally, the path coefficients to feedback practices were not stronger than those to feedback attitudes, which was counter to our expectations (see Figure 3).

Associations with two-factor study process and feedback self-efficacy

In the English and Turkish samples, we examined learning approaches using sections of the Revised Two-Factor Study Process Questionnaire (R-SPQ-2F; Biggs, Kember, and Leung 2001). In this model, deep motivation reflects genuine interest and critical engagement with material, while surface motivation represents minimal effort aimed at meeting requirements. We hypothesized positive associations between feedback literacy and deep motivation, as students with deep motivation are likely to view feedback as a tool for improvement. Conversely, we expected negative associations between feedback literacy and surface motivation, as students with a surface approach are less likely to value feedback. We further anticipated stronger links between these motivations and feedback practices than attitudes. Both deep and surface motivations were measured with five items each, rated on a 5-point Likert scale from (1) "strongly disagree" to (5) "strongly agree."

Feedback self-efficacy, assessed using the Feedback Orientation Scale (FOS; Linderbaum and Levy 2010), reflects confidence in using feedback effectively. While feedback self-efficacy is task-specific, feedback literacy encompasses broader knowledge, attitudes, and dispositions toward feedback. We hypothesized strong correlations between feedback self-efficacy and feedback literacy, as effective feedback use likely depends on literacy. Feedback self-efficacy was measured with five items (e.g. "I know that I can handle the feedback that I receive") on a 5-point Likert scale.
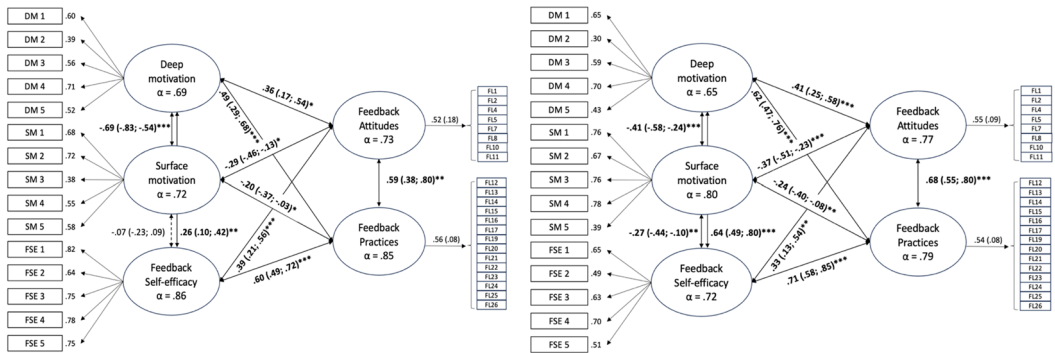
Figure 4. SEM model estimating associations between deep motivation, surface motivation, feedback self-efficacy, and feedback literacy dimensions in the English (left) and Turkish (right) samples. Mean factor loadings and standard deviation of feedback literacy indicators are shown to reduce visual clutter. Parentheses of path coefficients report 95% confidence intervals. ** $p < 0.01$; ** $p < 0.01$; *** $p < 0.001$.

Unlike the previous models, here, we did not assume a causal structure between these variables and feedback literacy, particularly feedback self-efficacy, as it operates on a similar level of abstraction, and we have no substantive rationale for causal assumptions. Instead, we modeled non-causal associations to assess convergent validity, accounting for measurement error using SEM. The SEM included deep motivation, surface motivation, feedback self-efficacy, feedback attitudes, and feedback practices, with all variables modeled as endogenous. Residual covariances were specified between items within facets for feedback literacy. The model fit adequately for the English sample: $X^2(541) = 856.81$, $p < 0.001$; $X^2/df = 1.58$; CFI = 0.88; TLI = 0.87; SRMR = 0.07; RMSEA = 0.05 (90% CI: 0.04–0.05). For the Turkish sample, model fit was excellent: $X^2(478) = 781.84$, $p < 0.001$; $X^2/df = 1.64$; CFI = 0.96; TLI = 0.96; SRMR = 0.06; RMSEA = 0.03 (90% CI: 0.00–0.04). As hypothesized, deep motivation was positively correlated with feedback literacy dimensions, showing stronger associations with feedback practices than attitudes, though with overlapping confidence intervals. Surface motivation was negatively correlated with feedback literacy dimensions. Feedback self-efficacy exhibited the strongest correlations with feedback literacy, reflecting its close conceptual alignment. Results for the Turkish sample mirrored those of the English sample, with slightly stronger associations and consistent patterns, suggesting stability of feedback literacy measurement and its relationships with other constructs across samples (see Figure 4).

Discussion

Psychometric properties of the SFLI across languages

The psychometric evaluation of the SFLI across three languages supported the two-factor model of feedback attitudes and practices (Woitt et al. 2025). While the German version required minimal adjustments, modifications were needed for the Turkish and English versions, such as removing certain items to achieve good fit indices. These variations may reflect linguistic and cultural differences in how students approach and interpret feedback. For instance, the item "When evaluating feedback, I take into account that it was shaped by the providers' perspective" (FL10) performed inconsistently across languages, which may suggest sociocultural nuances of feedback being interpreted differently based on educational and cultural backgrounds.

These findings align with research emphasizing the importance of examining feedback processes through an intercultural lens (e.g. Rovagnati and Pitt 2022; Pazio Rossiter and Bale 2023). For example, Pazio Rossiter and Bale (2023) showed that international students varied in their

interpretation of feedback, such as distinguishing suggestions from essential comments or balancing politeness with honest feedback. The modifications in the Turkish and English versions highlight that while feedback literacy is universally relevant, student engagement with feedback is shaped by local educational practices and values. More broadly, this underscores the importance of adapting educational constructs and measurement tools for diverse cultural contexts to ensure validity (Beaton et al. 2000).

Development of the short-form SFLI

A key objective of this study was to develop SFLI-S, a short-form version of the SFLI, retaining psychometric robustness while offering greater practicality for large-scale or time-constrained research. The 11-item SFLI-S demonstrated good fit across languages and preserved the core constructs of feedback attitudes and feedback practices. By reducing survey length by over 50% compared to existing feedback literacy instruments (Song 2022; Yildiz, Bozpolat, and Hazar 2022; Zhan 2022; Dawson et al. 2024; Woitt et al. 2025), the SFLI-S facilitates feedback literacy research in constrained contexts. Consistent factor loadings across samples support its use as a viable alternative to the full SFLI, especially in resource-limited settings. A further strength of the SFLI-S is its omission of item FL15, which has shown some differential item functioning between genders in a German sample.

However, some shortcomings should be noted. The reduction to one item per facet narrows the breadth of these facets, which may affect construct validity. Such trade-offs between brevity and validity are common in short-form instruments (e.g. Big Five short forms; Herzberg and Brähler 2006; Rammstedt and John 2007). Further research is needed to confirm that the SFLI-S adequately captures feedback literacy. Additionally, replacing items in the English and Turkish versions for the appraisal facet means that this facet is not represented by the conceptually strongest items. This issue was also seen in Woitt et al. (2025), where the appraisal facet unexpectedly loaded on the attitudes factor instead of the behavioral factor (initially feedback processing, later feedback practices).

Implications for research and practice

The SFLI provides a psychometrically sound instrument for assessing feedback literacy development in higher education. Each dimension of the model aligns with distinct potential intervention strategies: feedback attitudes can be cultivated by fostering a productive, agency-centered view of feedback, shifting students from passive reception to active engagement. Reflective activities, for example, help students rethink feedback's role and purpose. Feedback practices, in contrast, involve skills like processing feedback constructively, seeking it proactively, and adapting behaviors. Little et al. (2024) noted that most interventions rely on ad-hoc definitions of feedback literacy, lacking a consistent psychometric foundation. Our validated, multilingual instrument fills this gap, with full-length and short versions suited to diverse research and teaching needs. The SFLI and SFLI-S also enable comparison across studies, as Little et al. (2024) advocate, providing a foundation for cumulatively refining future instructional approaches.

Limitations

While this study provides strong initial evidence for the validity and reliability of the SFLI in German, English, and Turkish, some limitations warrant further investigation. First, the modifications to the Turkish and English versions suggest that certain items may not be fully equivalent across languages or reflect cultural differences in how feedback is perceived. Future research should explore which of these explanations applies and their implications.

We observed high correlations between feedback attitudes and practices, raising questions about their distinctiveness. Although our comparison with a single-factor model and conceptual reasoning supports treating them as distinct yet strongly linked—student beliefs likely shape their feedback engagement—further research is needed to clarify their boundaries. For instance, recent work (Weidlich et al. 2025) found that feedback attitudes interacted with feedback conditions to influence student motivation. A better understanding of what differentiates these dimensions could inform strategies to develop feedback literacy through targeted educational measures.

Conclusion

This study provides validity evidence for the Student Feedback Literacy Instrument (SFLI) across three languages and introduces a short-form version (SFLI-S) suitable for large-scale research and educational practice. The findings provide empirical support for the two-factor model we proposed in Woitt et al. (2025), affirming that student feedback literacy consists of both attitudinal and behavioral components. As feedback literacy research continues to proliferate, psychometrically sound measurement instruments are the foundation of rigorous quantitative research into the construct. We hope that the SFLI and SFLI-S contribute to this worthwhile goal.

Additional information

All SFLI items and Supplemental Materials are available at OSF.io: https://osf.io/e73jy/.

Disclosure statement

The authors report no conflicts of interest for this manuscript.

Ethics declaration

Informed consent for study participation as well as consent to publish the data in an academic paper was obtained from all study participants prior to collecting their data. Ethical approval for the second German sample was obtained through the institute's ethics board. For the Turkish sample, no ethical approval is necessary for survey studies with informed consent, given that data confidentiality and no disadvantage for non-participants is ensured. Participants recruited through panel data and online sampling services explicitly granted permission to usage of their data for research purposes.

ORCID

J. Weidlich http://orcid.org/0000-0002-1926-5127
I. Jivet http://orcid.org/0000-0002-8715-2642
D. Orhan Göksün http://orcid.org/0000-0003-0194-0451
J. Kraus http://orcid.org/0000-0001-7015-8477
H. Drachsler http://orcid.org/0000-0001-8407-5314

References

Beaton, D. E., C. Bombardier, F. Guillemin, and M. B. Ferraz. 2000. "Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures." *Spine* 25 (24): 3186–3191. doi:10.1097/00007632-200012150-00014.
Biggs, J., D. Kember, and D. Y. Leung. 2001. "The Revised Two-Factor Study Process Questionnaire: R-SPQ-2F." *British Journal of Educational Psychology* 71 (1): 133–149. doi:10.1348/000709901158433.

Bless, H., M. Wänke, G. Bohner, R. F. Fellhauer, and N. Schwarz. 1994. "Need for Cognition: Eine Skala Zur Erfassung Von Engagement Und Freude Bei Denkaufgaben." *Zeitschrift für Sozialpsychologie* 25 (2): 147–154.

Brown, T. A. 2015. *Confirmatory Factor Analysis for Applied Research*. New York: Guilford Publications.

Butler, D. L., and P. H. Winne. 1995. "Feedback and Self-Regulated Learning: A Theoretical Synthesis." *Review of Educational Research* 65 (3): 245–281. doi:10.3102/00346543065003245.

Cacioppo, J. T., R. E. Petty, J. A. Feinstein, and W. B. G. Jarvis. 1996. "Dispositional Differences in Cognitive Motivation: The Life and Times of Individuals Varying in Need for Cognition." *Psychological Bulletin* 119 (2): 197–253. doi:10.1037/0033-2909.119.2.197.

Carless, D., and D. Boud. 2018. "The Development of Student Feedback Literacy: Enabling Uptake of Feedback." *Assessment & Evaluation in Higher Education* 43 (8): 1315–1325. doi:10.1080/02602938.2018.1463354.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, NJ: Erlbaum.

Costello, A. B., and J. Osborne. 2005. "Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most from Your Analysis." *Practical Assessment, Research, and Evaluation* 10 (1): 7.

Dawson, P., Z. Yan, A. Lipnevich, J. Tai, D. Boud, and P. Mahoney. 2024. "Measuring What Learners Do in Feedback: The Feedback Literacy Behaviour Scale." *Assessment & Evaluation in Higher Education* 49 (3): 348–362. doi:10.1080/02602938.2023.2240983.

Dong, Z., Y. Gao, and C. D. Schunn. 2023. "Assessing Students' Peer Feedback Literacy in Writing: Scale Development and Validation." *Assessment & Evaluation in Higher Education* 48 (8): 1103–1118. doi:10.1080/02602938.2023.2175781.

Duckworth, A. L., C. Peterson, M. D. Matthews, and D. R. Kelly. 2007. "Grit: Perseverance and Passion for Long-Term Goals." *Journal of Personality and Social Psychology* 92 (6): 1087–1101. doi:10.1037/0022-3514.92.6.1087.

Duckworth, A. L., and P. D. Quinn. 2009. "Development and Validation of the Short Grit Scale (GRIT–S)." *Journal of Personality Assessment* 91 (2): 166–174. doi:10.1080/00223890802634290.

Hagger, M. S., and K. Hamilton. 2019. "Grit and Self-Discipline as Predictors of Effort and Academic Attainment." *The British Journal of Educational Psychology* 89 (2): 324–342. doi:10.1111/bjep.12241.

Hattie, J., and H. Timperley. 2007. "The Power of Feedback." *Review of Educational Research* 77 (1): 81–112. doi:10.3102/003465430298487.

Herzberg, P. Y., and E. Brähler. 2006. "Assessing the Big-Five Personality Domains via Short Forms." *European Journal of Psychological Assessment* 22 (3): 139–148. doi:10.1027/1015-5759.22.3.139.

Hoebel, J., S. Müters, B. Kuntz, C. Lange, and T. Lampert. 2015. "Measuring Subjective Social Status in Health Research with a German Version of the MacArthur Scale." *Bundesgesundheitsblatt, Gesundheitsforschung, Gesundheitsschutz* 58 (7): 749–757. doi:10.1007/s00103-015-2166-x.

Howard, M. C. 2016. "A Review of Exploratory Factor Analysis Decisions and Overview of Current Practices: What we Are Doing and How Can we Improve?" *International Journal of Human-Computer Interaction* 32 (1): 51–62. doi:10.1080/10447318.2015.1087664.

Hu, L. T., and P. M. Bentler. 1999. "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives." *Structural Equation Modeling: A Multidisciplinary Journal* 6 (1): 1–55. doi:10.1080/10705519909540118.

International Test Commission. 2017. *The ITC Guidelines for Translating and Adapting Tests*. 2nd ed. https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf

Kraus, J. M., Y. Forster, S. Hergeth, and M. Baumann. 2019. "Two Routes to Trust Calibration: Effects of Reliability and Brand Information on Trust in Automation." *International Journal of Mobile Human Computer Interaction)*11 (3): 1–17. doi:10.4018/IJMHCI.2019070101.

Lam, K. K. L., and M. Zhou. 2022. "Grit and Academic Achievement: A Comparative Cross-Cultural Meta-Analysis." *Journal of Educational Psychology* 114 (3): 597–621. doi:10.1037/edu0000699.

Linderbaum, B. A., and P. E. Levy. 2010. "The Development and Validation of the Feedback Orientation Scale (FOS)." *Journal of Management* 36 (6): 1372–1405. doi:10.1177/0149206310373145.

Little, T., P. Dawson, D. Boud, and J. Tai. 2024. "Can Students' Feedback Literacy be Improved? A Scoping Review of Interventions." *Assessment & Evaluation in Higher Education* 49 (1): 39–52. doi:10.1080/02602938.2023.2177613.

Liu, Q., and J. C. Nesbit. 2024. "The Relation between Need for Cognition and Academic Achievement: A Meta-Analysis." *Review of Educational Research* 94 (2): 155–192. doi:10.3102/00346543231160474.

MachineTranslation.com. 2023. *DeepL vs Google Translate: Which is better for machine translation?* Retrieved September 1, 2024, from https://www.machinetranslation.com/blog/deepl-vs-google-translate

Molloy, E., D. Boud, and M. Henderson. 2020. "Developing a Learning-Centred Framework for Feedback Literacy." *Assessment & Evaluation in Higher Education* 45 (4): 527–540. doi:10.1080/02602938.2019.1667955.

Pazio Rossiter, M., and R. Bale. 2023. "Cultural and Linguistic Dimensions of Feedback: A Model of Intercultural Feedback Literacy." *Innovations in Education and Teaching International* 60 (3): 368–378. doi:10.1080/14703297.2023.2175017.

Porter, S. R., M. E. Whitcomb, and W. H. Weitzer. 2004. "Multiple Surveys of Students and Survey Fatigue." *New Directions for Institutional Research* 2004 (121): 63–73. doi:10.1002/ir.101.

Rammstedt, B., and O. P. John. 2007. "Measuring Personality in One Minute or Less: A 10-Item Short Version of the Big Five Inventory in English and German." *Journal of Research in Personality* 41 (1): 203–212. doi:10.1016/j.jrp.2006.02.001.

Rolstad, S., J. Adler, and A. Rydén. 2011. "Response Burden and Questionnaire Length: Is Shorter Better? A Review and Meta-Analysis." *Value in Health: The Journal of the International Society for Pharmacoeconomics and Outcomes Research* 14 (8): 1101–1108. doi:10.1016/j.jval.2011.06.003.

Rovagnati, V., and E. Pitt. 2022. "Exploring Intercultural Dialogic Interactions between Individuals with Diverse Feedback Literacies." *Assessment & Evaluation in Higher Education* 47 (7): 1057–1070. doi:10.1080/02602938.2021.2006601.

Schmidt, F. T., J. Fleckenstein, J. Retelsdorf, L. Eskreis-Winkler, and J. Möller. 2019. "Measuring Grit: A German Validation and a Domain-Specific Approach to Grit." *European Journal of Psychological Assessment* 35 (3): 436–447. doi:10.1027/1015-5759/a000407.

Schreiber, J. B. 2008. "Core Reporting Practices in Structural Equation Modeling." *Research in Social & Administrative Pharmacy: RSAP* 4 (2): 83–97. doi:10.1016/j.sapharm.2007.04.003.

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. doi: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. doi:10.1080/14703297.2012.647781.

Weidlich, J., A. Fink, I. Jivet, S. Gombert, T. Giorgiashvili, J. Yau, and H. Drachsler. (2025). "Highly informative feedback using learning analytics: how feedback literacy moderates student perceptions of feedback." Manuscript submitted for publication.

Winstone, N. E., G. Mathlin, and R. A. Nash. 2019. "Building Feedback Literacy: Students' Perceptions of the Developing Engagement with Feedback Toolkit." *Frontiers in Education* 4: 39. doi:10.3389/feduc.2019.00039.

Wisniewski, B., K. Zierer, and J. Hattie. 2019. "The Power of Feedback Revisited: A Meta-Analysis of Educational Feedback Research." *Frontiers in Psychology* 10: 3087. doi:10.3389/fpsyg.2019.03087.

Woitt, S., J. Weidlich, I. Jivet, D. Orhan Göksün, H. Drachsler, and M. Kalz. 2025. "Students' Feedback Literacy in Higher Education: An Initial Scale Validation Study." *Teaching in Higher Education* 30 (1): 257–276. doi:10.1080/13562517.2023.2263838.

Yildiz, H., E. Bozpolat, and E. Hazar. 2022. "Feedback Literacy Scale: A Study of Validation and Reliability." *International Journal of Eurasian Education and Culture* 7 (19): 2214–2249. doi:10.35826/ijoecc.624.

Yu, S., E. Di Zhang, and C. Liu. 2022. "Assessing L2 Student Writing Feedback Literacy: A Scale Development and Validation Study." *Assessing Writing* 53: 100643. doi: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. doi:10.1080/02602938.2021.2001430.