Technology, Knowledge and Learning
https://doi.org/10.1007/s10758-026-09972-9

ORIGINAL RESEARCH

Validity evidence for the internal structure of the peer feedback orientation scale

Marco Kalz[1] · **Annette Schulze**[1] · **Holger Meeh**[1] · **Joshua Weidlich**[2,3]

Received: 27 February 2025 / Revised: 7 March 2026 / Accepted: 10 March 2026

Abstract
Peer feedback has the potential to contribute to a transformation of the feedback culture in higher education. The disposition of students for peer feedback is a crucial factor for the successful implementation of peer feedback. In earlier work we have developed through a mixed-method sequential research approach an initial instrument to measure the peer feedback orientation of students. The current study has the goal produce further evidence for the validity of the factor structure of the initial instrument, compare competing models and to cross-validate the model across samples and sub-populations. For this purpose, the Peer feedback Orientation instrument (PFOS) has been implemented in a mandatory course on media education in the teacher education program of the institution of the first author. 269 students have participated in the study. A confirmatory factor analysis has confirmed the measurement model from an earlier exploratory factor analysis. A competing two-factor model focusing on the role-change in peer feedback processes could not achieve comparable benchmarks. A test of measurement invariance and a retesting of this measurement provides further evidence of the validity of the internal structure. These results show that the peer feedback orientation scale is a robust measure for students in higher education and that peer feedback orientation is a promising area for future research. By providing an empirically supported measurement model, the study offers a theoretically grounded framework for examining how students' orientations toward peer feedback shape their participation in feedback activities. This framework may help to explain variability in peer-feedback engagement and outcomes observed in higher education settings and thus contributes to the broader effort to better understand the learner dispositions that support effective feedback processes.

Keywords Peer feedback · Higher education · Peer feedback orientation · Dispositions · Confirmatory factor analysis · Individual differences · Instrument development

Extended author information available on the last page of the article

Published online: 06 April 2026

1 Introduction

Topping (2009) defines peer assessment as "an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners" (p. 250). Similarly, peer feedback can be understood as both a source of feedback and a learning activity in which students provide comments to one another on their work (Liu & Carless, 2006). A main barrier to expanding feedback opportunities in higher education has been identified as the workload connected to elaborated individual feedback from educators (Henderson et al., 2019; Wild & Esdar, 2014). Therefore, peer feedback has been recognized as a scalable practice with potential to be used in multiple educational contexts beyond fully-online courses (Double et al., 2020; Authors, 2021; Rodríguez et al., 2022).

Despite documented benefits (see e.g. Nicol et al., 2014; Tai et al., 2018), peer feedback has not yet been widely adopted in higher education. One reason for the lack of large-scale implementation is the implementation complexity of peer feedback in designing and implementing peer feedback activities. For example, Topping (1998) identified 17 design dimensions that instructors must consider when implementing peer assessment, including subject area, directionality of feedback, voluntariness, ability level of participants, and reward structures. Subsequent work has further expanded and refined these design considerations (e.g., Gielen et al., 2011; Topping, 2017). Although the use of digital tools has reduced some logistical barriers, implementing peer feedback still requires navigating numerous pedagogical design decisions (Kerman et al., 2024).

1.1 Dispositions Toward Peer Feedback

An important determinant of whether peer feedback succeeds is students' attitudes and beliefs about the practice (Greisel et al., 2025). Consequently, some peer-feedback research has focused on how students are disposed toward peer feedback (see Table 1 for an overview), albeit with varying emphases and theoretical backgrounds. For example, Wen and Tsai (2006) conceptualized peer-feedback as an attitudinal challenge towards online-feedback and developed an instrument with four subscales, whereas Huisman et al., (2020) approached peer-feedback from the perspective of beliefs, also identifying four dimensions. Friedman et al., (2008) instead built their work on motivation and expectancy theory.

During the development process of our instrument, more research on *general* feedback literacy has been published (see Nieminen & Carless, 2022). But we have assumed that existing approaches to general feedback literacy might only have limited validity for the context of peer feedback because of the role-change involved during the peer-feedback process. Existing conceptualizations of feedback literacy were largely developed in the context of teacher–student feedback and therefore focus primarily on students' capacities to interpret and use feedback as recipients, albeit active and engaged recipients (Carless & Boud, 2018). However, peer feedback requires learners to additionally act as feedback providers and therefore involves additional attitudes and skills. This was also the motivation by Dong et al. (2023) who have developed a peer-feedback literacy scale focused on writing contexts (see Table 1), including dimensions like willingness to participate or cooperative Learning Ability, which they distinguish into cognitive aspects of readiness and social-affective readiness for peer feedback.

Table 1 Dispositions toward peer-feedback overview

Theoretical rooting	Main assumptions	Authors	Validated instruments/ Scales and Subscales
Attitudes	Peer-feedback requires an attitudinal change in students	Wen and Tsai (2006)	Positive Attitude Subscale (POS), Online Attitude Subscale (OAS), Understanding-and-Action Subscale (UAS) and Negative Attitude Subscale (NAS)
Motivation & expectancy theory	Peer-feedback acceptance depends on valence and force	Friedman et al. (2008)	Team functioning, Effort and assessments fairness, Team concerns
Beliefs & values	Peer-feedback success relies on beliefs and values of student about the method and the quality of feedback	Huisman et al. (2020)	Valuation of peer feedback as instructional method, confidence in quality of received peer feedback, confidence in own peer feedback quality, valuation of peer feedback as an important skill
Peer-feedback literacy	Peer-feedback is a specific ability	Dong et al. (2023)	Feedback-related knowledge and abilities (FKA), Willingness to participate (WP), Appreciation of peer feedback (APF) and cooperative learning ability (CLA). The authors summarize those factors as cognitive readiness (FKA & CLA) and social-affective readiness for peer feedback (WP and APF)

Our review of the literature suggests that theory-building on the acceptance and success of peer feedback remains limited. Existing studies tend to focus on isolated aspects, such as beliefs or motivation, or examine peer feedback in highly specific contexts, most commonly writing tasks. We found little work that adopts a holistic measurement approach capable of capturing the complexity of peer feedback activities, particularly the interplay between providing and receiving feedback and between domain- and task-related factors. This limited understanding of students' dispositions toward peer feedback also makes it difficult to translate the many design dimensions identified in the literature into actionable implementation decisions. To address this challenge, we focus on the concept of *peer-feedback orientation* as a theory-informed lens for reducing the complexity of peer feedback design.

1.2 Peer Feedback Orientation

London and Smither (2002) have defined feedback orientation as "an individual's overall receptivity to feedback, including comfort with feedback, tendency to seek feedback and

process it mindfully, and the likelihood of acting on the feedback to guide behavior change and performance improvement" (p. 1).

Extending this we define peer-feedback orientation as a multidimensional disposition reflecting learners' motivational, social, and communicative readiness to participate in reciprocal feedback processes with peers and engaging productively with peer feedback. It captures how students value peer feedback, how receptive and accountable they feel toward it, how confident they are in dealing with it, and how ready they are to participate in the communicative exchanges peer feedback entails. In this sense, peer feedback orientation is distinct from peer feedback skills: it does not measure what students can do in a performance sense, but rather the dispositional conditions that may support or hamper effective participation in peer feedback processes.

Further, peer feedback orientation overlaps with, but is not identical to, peer feedback literacy. Feedback literacy frameworks typically emphasize the capacities needed to appreciate, interpret, and use feedback effectively (Carless & Boud, 2018; Molloy et al., 2020; Woitt et al., 2025). By contrast, orientation refers to a broader dispositional stance toward participation in peer-feedback situations, including perceived value, responsibility, receptivity, confidence, and communicative readiness. The present instrument therefore aims not to replace literacy-oriented approaches, but to capture a related construct that may help explain why students differ in their willingness and readiness to engage in peer feedback even before performance in such tasks is observed.

Based on the lack of relevant work in education research, we identified promising work on feedback orientation from the business context. For their feedback orientation scale (FOS), Linderbaum and Levy (2010) argued that identifying and adapting to individual differences in feedback is important for the business context and that this focus on individual differences will enable organisations to make better use of feedback. To this end, building on the Theory of Planned Behavior (Ajzen, 1991), they developed a multidimensional self-report instrument of feedback orientation able to measure individual differences for the purpose of enhancing feedback uptake. The approach included exploratory (EFA) and confirmatory factor analysis (CFA) yielding a four-factor solution: utility, accountability, social awareness and feedback self-efficacy.

In the FOS (Linderbaum & Levy, 2010), these are understood as follows:

- Utility refers to the belief that feedback serves a useful purpose in helping a person achieve goals or desired outcomes. It captures whether feedback is seen as meaningful, instrumental, and relevant for improving performance or personal effectiveness.
- Accountability describes the tendency to feel responsible for responding to feedback and taking action based on it. It reflects the sense that feedback should lead to follow-up rather than remain without consequences.
- Social awareness refers to the tendency to use feedback to understand how one is perceived by others and to stay attentive to those perceptions. It emphasizes sensitivity to external perspectives and the use of feedback to adjust one's behavior accordingly.
- Feedback self-efficacy denotes a person's confidence in their ability to understand feedback correctly and respond to it constructively. It reflects the belief that one has the competence and control needed to deal effectively with feedback situations.

Considering the focus of the instrument we deemed this a promising starting point for our research. Our research goal was to extend this scale and adapt it to the peer feedback context. Consequently, we first invested in ensuring that the context change from the business context to the higher education context is supported by expert judgements and then we deducted items for a theory-based and data-based analysis of the scales' validity.

1.3 Development of the Peer Feedback Orientation Instrument

To transfer the instrument from the business context to the higher education context, we conducted an initial qualitative study (Kasch et al., 2023). Semi-structured interviews conducted with students, instructors, and researchers (n = 13) identified a wide spectrum of individual factors shaping their orientation toward peer feedback. Thematic analysis indicated that the most salient influences concerned the perceived value of both giving and receiving peer feedback, the quality of interpersonal relationships among students, perceptions of fairness, and relevant competencies. Importantly, the relevance of established feedback orientation dimensions (utility, accountability, social awareness, and self-efficacy) was corroborated within the higher education context. At the same time, the findings revealed context-specific interpretations of these dimensions, suggesting the need to design a peer-feedback orientation scale tailored explicitly to higher education settings.

In a follow-up study (Kasch et al., 2022) we built on this initial work by generating self-report items and identifying an initial factor structure via EFA on a sample of 148 higher education students. We identified a correlated-factor model with five factors describing the peer feedback orientation of students: accountability, communicativeness, utility, self-efficacy and receptivity. For those five factors, a simple structure was identified: 4 factors are moderately correlated (accountability, utility, self-efficacy, receptivity), while one factor has been identified to be independent (communicativeness)—an extension to the initial work by Linderbaum and Levy (2010) and Authors (2021). Communicativeness as a separate factor in peer feedback settings makes conceptual sense, because it is not only a matter of interpreting information, but also of expressing critique, asking for clarification, and engaging in sustained dialogue with peers. We therefore interpret communicativeness not as a data-driven residual factor, but as a context-specific extension that reflects the dialogic nature of peer feedback (Ajjawi & Boud, 2017; Zhu & Carless, 2018). The final instrument consisted of a 22-item questionnaire with sound psychometric properties, including a clear factor structure and good internal consistency.

1.4 Research Questions

Building on Kasch et al. (2021) and Kasch et al. (2023), the purpose of the current study is to further develop both theory and measurement of students' peer feedback orientation. To this end, our goal is to provide further validity evidence for the PFOS dimensionality and psychometric functioning. Evidence concerning the internal structure of an instrument attests to validity by corroborating theorized constructs with empirically observed factors. Given the multidimensional nature of peer feedback orientation, examining the internal structure is therefore a critical step toward evaluating whether the proposed dimensions cohere in theoretically meaningful ways (Reeves & Marbach-Ad, 2016). The present study focuses on this aspect of validity by testing the fit and stability of the previously identified factor structure

through structural equation modeling. The current study seeks to validate the measurement model, previously developed through an exploratory factor analysis (EFA), through a new confirmatory factor analysis (CFA) and to compare it to a competing model based on the role-change in peer feedback activities. Furthermore, a cross-validation of the prioritized model should be conducted. These research needs were transferred into the following three research questions:

RQ1: Can we confirm the initial factor structure from the EFA and the specific extension of the feedback orientation scale with a fifth factor for the peer feedback context?

RQ2: How well does a two-factor solution perform in comparison which divides the peer feedback process in the roles as peer feedback receiver and peer feedback provider?

RQ3: Does the prioritized model show its robustness across sub-groups and samples?

2 Methods

The study design follows a survey-based approach to collect data about the dispositions of students regarding peer feedback. We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study (Simmons et al., 2012). The sample-size of this study followed prior recommendations by Kyriazos (2018a). The author discusses the challenge to define an adequate sample size in CFA studies. He recommends large sample sizes if there are a low number of indicators per factor and if there are a very high number of factors. The indicator-to-factor ratio in the instrument used for the data collection follows the recommended number of at least three indicators per factor, while most factors have 4 or 5 indicators. Furthermore, 5 factors is not seen as a very high number of factors according to the recommendations. Since we are building on results of a prior exploratory factor analysis we assume that we have identified a "simple structure" before and set a requirement of a ratio of a minimum of 10 data points per item as the ideal sample size which is regarded as a common rule of thumb. This would require at least 220 participants in the current study. Simulation studies have shown that a sample size of more than 200 combined with an item-factor ratio of at least three has led to no convergence failures and no improper solutions (Kyriazos, 2018a). We build on recommendations by Jackson et al. (2009) and Kyriazos (2018b) for the reporting of our study, its methodology and results.

2.1 Participants

Participants of this study were recruited in the context of a mandatory module on media education in teacher education programs at [blinded for peer review]. In 12 concurrent courses on foundations of media education a questionnaire was implemented with a total amount of approximately 350 students. From these students, 269 filled in the questionnaire fully while 54 participants filled it out partially. Only full datasets were further analyzed leading to absence of missing data. The mean age of participants was 22.02 years (SD = 2.28). 199 participants were female, while 67 participants male, 2 participants selected a diverse gender category while 1 participant chose to not respond to this question. We have applied a social interpretation of the gender category according to the guidelines by Döring (2013) and Heidari et al. (2016). This gender distribution is typical for a study in a teacher education program (UNESCO, 2023). The questionnaire was placed before a peer feedback activ-

ity. Participation in the research was voluntary and all participants agreed to an informed consent declaration online before participation. No further incentives were set for participation in the research.

2.2 Materials

The questionnaire implemented in this study consists of some basic demographic information and the same items used in our prior exploratory factor analysis study. The research followed a sequential-mixed method paradigm in which an existing scale from the business context (Linderbaum & Levy, 2010) was adapted to the context of higher education [blinded for peer review]. An exploratory factor analysis of those items resulted in a 22-item scale to measure the peer feedback disposition of students which had an underlying structure of 5 factors [blinded for peer review]. These original items were translated from Dutch into German and retranslated into Dutch to ensure the same meaning. A five-point-Likert scale was used ranging from "totally disagree" to "totally agree". The instrument was implemented in the Limesurvey system for online surveys. The papaya-package was used for the writing of this manuscript. The lavaan package was used for the confirmatory factor analysis (CFA) (Rosseel, 2012).

2.3 Procedure

Students filled out the questionnaire as preparation of a peer feedback activity.

Participation in the survey was voluntary and an informed consent and declaration of dataprotection and data-security aspects of the study was delivered before participants started the survey. Data collection was conducted at the institution of the first author from May until July 2023 in a mandatory module on media education for all teacher education programs at the institution. The link to the online questionnaire was integrated in the digital learning environment. The data analysis was conducted from August until October 2023. As an initial step, the measurement model was constructed based on the results of the exploratory factor analysis described in (Kasch et al., 2022) (Model 1). Furthermore, we tested an alternative model consisting of a two-factor solution (receiving and providing feedback, Model 2). After controlling fit indices and testing potential adaptations of each model we selected the best fitting model and further tested this model for measurement invariance and controlled if we can confirm the model with an additional dataset (test–retest reliability).

2.4 Data Analysis

We used R (Version 4.3.2; R Core Team, 2023) and the R-packages papaja (Version 0.1.1; Aust & Barth, 2022), and tinylabels (Version 0.2.3; Barth, 2022) for all our analyses. Confirmatory factor analysis was conducted according to guidelines by Jackson et al. (2009), Tabachnick et al. (2013) and finally Kyriazos and Poga-Kyriazou (2023). After finalisation of the data collection, we first controlled the data for completeness. 13 outliers were identified with the interquartile range criterion (IQR) and deleted according to guidelines by Tabachnick et al. (2013). The IQR criterion and a cut-off point of 1.5 has been proposed by Tukey (1977) and further refined and evaluated by other researchers as a robust method for outlier detection (Hoaglin & Iglewicz, 1987). The data was tested for multicollinearity

based on inter-item correlations of all indicators and we checked Variance Inflation Factors (VIF). Factor correlations of latent constructs was checked following guidelines by Kyriazos and Poga (2023). After deleting outliers, we checked the ratio between items and participants and controlled for multivariate normal distribution of the data. As the next step, we assessed whether the data are suited for factor analysis. We then conducted a confirmatory factor analysis with the Lavaan package in R (Rosseel, 2012) and evaluated the model based on standard threshold values for factor analysis. Following recommendations mentioned above, we combined multiple fit indices to evaluate the fit of the models to the data. No item parceling was used for the analysis.

3 Results

3.1 Data-Screening and Sample

Only complete data were retained for analysis (n = 269). 13 outliers were deleted from the dataset resulting in 256 cases included in the further analysis following the interquartile range approach. Since the PFOS instrument consists of 22 items in total, the item-participant ratio was 11.63 data points per item which is considered as acceptable standard for confirmatory factor analysis. A Bartlett's Test of Sphericity showed that the data is suited for factor analysis with a highly significant test result. This result was confirmed through a check of the Kaiser-Mayer-Olkin criterion which also confirmed factorability of the data (KMO = 0.77).

3.2 Univariate and Multivariate Normality

The MVN package in R was used to control for multivariate normal distribution of the data. The Mardia's test for multivariate normality showed that the data suffers from skewness and kurtosis and does not follow a normal distribution. The data showed substantial deviations from multivariate normality (multivariate kurtosis = 4111.42, z = 27.63, $p < .001$). Accordingly, a robust method was applied to conduct CFA on the data. Due to the use of Likert items in the PFOS instrument, we followed the latest guidelines by Li (2016) and Kyriazos and Poga-Kyriazou (2023) and used the Diagonally weighted least square (DWLS) estimator for the analysis.

3.3 Confirmatory Factor Analysis

Prior to estimating the CFA model, we examined multicollinearity among the observed indicators. Inter-item correlations were below the recommended threshold of .85, VIF values ranged from 1.04 to 1.66 indicating no problematic multicollinearity. At the latent level, factor correlations ranged from 0.06 to 0.51 and none exceeded .85. Together, these diagnostics indicate the absence of multicollinearity among the latent constructs.

To answer our first research question we have analysed our theoretical model and tested competing models. We used the conservative benchmarks formulated by Kyriazos (2018b) to evaluate the models. These benchmarks are depicted in the first line of Table 1. The initial confirmatory factor analysis of the measurement model which was the result of the EFA-

study by [blinded for peer review] could be estimated and was overidentified (Model 1, Fig. 1). The initial model showed very good fit measures.

Due to the fact that peer feedback disposition can also be modeled as a disposition for providing feedback and receiving feedback we also tested a competing two-factor model in which the indicators have been assigned to those two dimensions (Model 2). Model 2 was also overidentified but did not perform well on several fit measures and indices.

3.4 Measurement Invariance

While we could show that the theoretical model shows a good fit to the data, the question arises if the model is also robust across different groups. For the assessment of measurement invariance of the scale we follow guidelines by Putnick and Bornstein (2016) and Luong and Flake (2022). In terms of grouping variables that could be used to test measurement invariance, only the variable "Gender" was available. Other variables like age or course context either have too low variance for testing of measurement invariance or have too small samples. Also, the gender variable was not ideal for testing measurement invariance (due

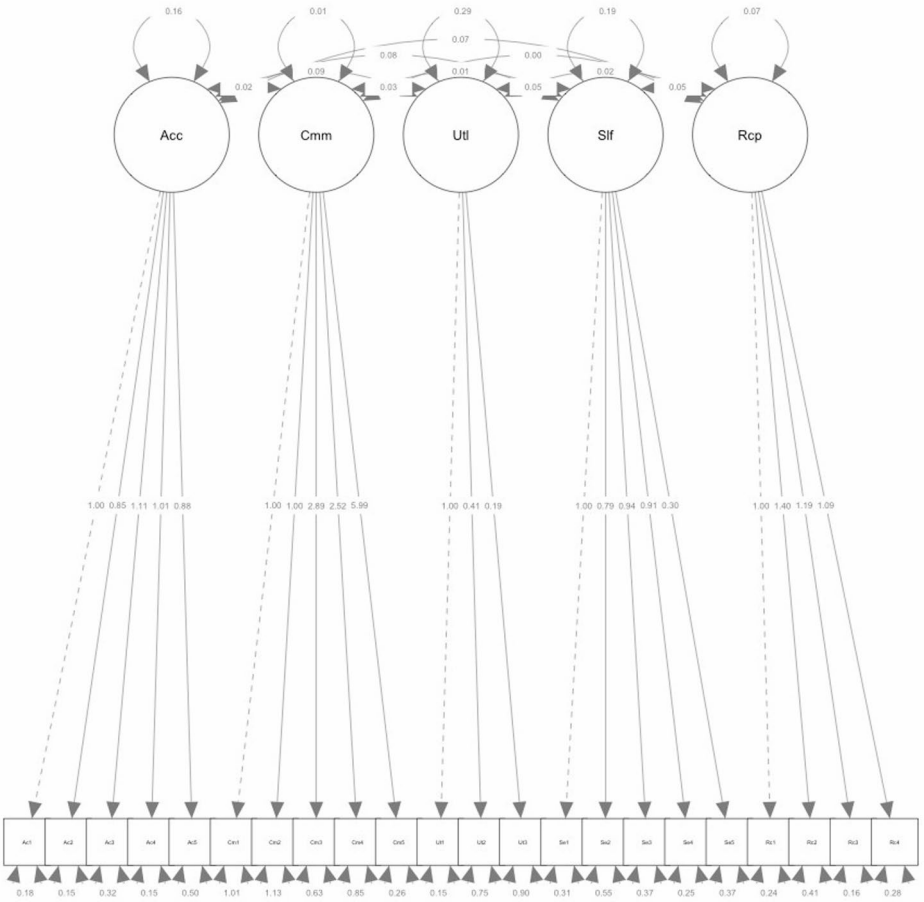

Fig. 1 CFA model 1 for peer feedback orientation of students

Table 2 Model fit of four CFA-models

Model	×2	df	x^2/df	*p*	TFI	CFI	RMSEA	SRMR
Benchmark	–	–	<3	>.05	>0.95	>0.95	<0.06	<0.08
Model 1	240.5	199	1.2	0.024	0.961	0.966	0.029	0.066
Model 2	322.73	208	1.55	0	0.896	0.906	0.048	0.082

Table 3 Measurement invariance

Model	x^2(df)	CFI	RMSEA (90% CI)	SRMR	Model comp	Δ CFI	Δ RMSEA	Δ SRMR	Decision
M1: CI	392.39 (396)	1	0.000 [0.0000;031]	0.077	–	–	–	–	–
M2: MI	471.34 (413)	0.954	0.34 [0.0130; 048]	0.084	M1	−0.046	0.034	0.007	Reject
M3: SI	493.44 (430)	0.950	0.35 [0.0160;049]	0.085	M2	−0.004	0.001	0.001	Reject

to only 67 male participants), but we tried the analysis with this variable. The question was thus if the scale was also invariant across male and female students. For the initial screening, we assessed the configural invariance (CI). This basic analysis answers the question if the relation between indicators and factors is valid across different groups. Configural non-invariance could suggest a problem in the theoretical model and would require further investigation into why constructs and indicators are perceived differently across groups. The next level of analysis we applied was the metric invariance (MI) which is an assessment of equality of factor loading across groups. Furthermore, we tested scalar invariance (SI), which represents that factor loadings and intercepts are equal across groups. Results of the model test are available in Table 2. We can see that only configural invariance can be confirmed with the available data.

3.5 Cross-validation

To examine whether the selected model could be reproduced in a different data collection, we estimated the same CFA model in the post-questionnaire sample with 197 participants. We again controlled for outliers and factorability of the data which could be confirmed. After deleting 23 outliers with the same approach as above, we have conducted a confirmatory factor analysis with the remaining 174 data points. For the cross-validation model 1 could be overidentified and achieved the expected benchmarks with Chi-square = 229.33, Chi-square/df = 1.15, RMSEA = 0.03, CFI = 0.98 and TLI = 0.92, SRMR = 0.078 (Table 3).

4 Discussion

During the initial study by [blinded for peer review], an existing instrument from the business context was adapted to the higher education context in the Dutch Higher Education system and a number of items were generated to assess the peer feedback orientation of students. In a further study [blinded for peer review], an exploratory factor analysis resulted in an initial 5-factor structure with the factors accountability, communicativeness, utility, self-

efficacy and receptivity. A simple structure could be identified: 4 factors were moderately correlated (accountability, utility, self-efficacy, receptivity), while one factor has been identified to be independent (communicativeness). The final instrument consisted of a 22-item questionnaire with satisfactory psychometric qualities. The current study was conducted to test this model and competing models and to cross-validate the prioritised model.

Due to the fact that the measurement model from the earlier EFA-analysis showed a very good performance and fit to the data, research question 1 (RQ1) could be answered affirmatively. The five identified factors show some similarities to earlier described general and specific (peer-)feedback scales. The identified factors are similar to the factors described by King et al. (2009) in the feedback orientation scale, namely utility, self-efficacy and communication competence. Furthermore, the factor receptivity is similar to the factor appreciation identified in the general feedback literacy scale by Zhan (2022) and self-efficacy is identical to the factor identified by Song (2022). Unsurprisingly there is overlap with the original scale by Linderbaum and Levy (2010) which has been taken as a starting point for the development of the peer feedback orientation scale. The underlying concept of "orientation" as a mix of skills and capabilities, attitudes and perception is the reason why the peer feedback orientation scale consists of a very different factor structure compared to the work by Wen and Tsai (2006). Their scale focuses solely on attitudes and actions and does not take into account skills or prior experiences. Furthermore, the "Beliefs about Peer feedback Questionnaire" by Huisman et al. (2020) situated peer feedback ability only in the belief system by students. Last but not least, the peer feedback orientation scale as a general instrument to measure and develop the peer feedback orientation of students is also different to the work by Dong et al. (2023) who have presented a scale for peer feedback literacy in writing. Only the factor receptivity of our instrument is similar to the factor "appreciation" of this scale.

To answer research question 2 (RQ2), a competing model has been tested: Peer feedback is often compared to other feedback activities in terms of the role switch between peer feedback receiver and peer-feedback provider which leads to specific procedural and cognitive requirements (Heine & Vesterager, 2024; Panadero, 2016). The two-factor model has divided the items of the scale into two factors: a factor related to the role of peer feedback receiver and a second factor related to the role of peer feedback provider. This simple two-factor model representing the dimensions of "providing feedback" and "receiving feedback" has been tested but the model did not perform well on the data. This answers the second research question.

To go beyond the model assessment, we also had the goal to analyse some psychometric properties of the prioritized model. These properties are construct validity and absence of bias. To analyse if the identified model is valid across groups in the sample and has not been resulted from systematic bias in the data we have applied measurement invariance and an analysis of test–retest reliability. The invariance analyses provide preliminary evidence of configural invariance across gender, suggesting that the overall pattern of relations between items and factors is similar in male and female students. However, because metric and scalar invariance were not supported, the present data do not justify stronger claims about equivalence of factor loadings, intercepts, or latent mean comparability across gender. These findings should therefore be interpreted as evidence of structural similarity rather than full cross-group comparability. Converging evidence stemming from a cross validation with a separate data collection allowed us to ascertain the reproducibility of the factor structure,

thereby further supporting the general stability of the model. With these results, the third research question is addressed (RQ3).

The study and its approach has several limitations. First, the convenience sample applied in the research is not the ideal scenario for research on validity of latent constructs. The chosen sample was from a teacher education program in which the specific traits of the participants and the research context may have influenced some of the results. Future work should apply more controlled levels of sampling. The collected data was not normally distributed leading to a specific approach for confirmatory factors analysis which fits to the data. While the sample-size for the study was not problematic in terms of the item-participant ratio or other recommendations for conducting CFA, the sample size was problematic to test for higher levels of measurement invariance. Only the gender variable could serve here as a grouping variable. With a larger sample, more options for measuring invariance would have been available.

5 Conclusion

In this paper, we provide further validity evidence for the factorial structure of the peer feedback orientation measure for higher education students. This finding supports the use of this instrument, leading to potential practical implications that include the use of adaptive designs aligned to each student's peer feedback orientation and further investigation of the construct as moderating factor in students feedback processes.

The present study contributes to peer feedback research by offering a theoretically grounded conceptualization of peer-feedback orientation as a multidimensional learner disposition. The confirmed five-factor structure suggests that students' engagement with peer feedback is shaped by a combination of motivational beliefs, perceived capabilities, and social-interactional orientations. This finding extends existing feedback research, which has often focused primarily on cognitive aspects of feedback use or on general feedback attitudes. In particular, the results highlight the inherently social nature of peer-feedback processes. While the factors utility and feedback self-efficacy capture learners' perceptions of the learning value of peer feedback and their capability to participate in such activities, the dimensions social awareness, communicativeness, and accountability emphasize the interpersonal and normative aspects of peer-feedback interactions. These dimensions reflect students' awareness of the social consequences of feedback, their willingness to engage in feedback dialogue, and their perceived responsibility toward peers. Together, these findings support the view that effective participation in peer feedback requires not only cognitive engagement but also social and communicative dispositions.

By providing an empirically supported measurement model, the study offers a theoretically grounded framework for examining how students' orientations toward peer feedback shape their participation in feedback activities. This framework may help to explain variability in peer-feedback engagement and outcomes observed in higher education settings and thus contributes to the broader effort to better understand the learner dispositions that support effective feedback processes.

While this study provides evidence for the internal structure of the PFOS, it represents only one component of a broader validity argument. A comprehensive validation requires integrating multiple categories of evidence, including content relevance, response processes,

relations to other variables, and the consequences of score use (Kane, 2006; Messick, 1995). Future research should therefore extend the validity argument by examining how PFOS scores relate to external criteria, how respondents interpret and engage with items, and how the instrument performs across diverse educational contexts. Such additional evidence would further strengthen the theoretical and practical foundations of the scale.

Several avenues for future research follow from the present findings. First, further studies should examine how the identified dimensions of peer-feedback orientation relate to actual feedback behaviors and learning outcomes. For instance, future research could investigate whether students with higher levels of communicativeness and social awareness provide more elaborated or constructive feedback, or whether feedback self-efficacy predicts students' willingness to revise their work based on peer comments.

Second, future research could explore how peer-feedback orientation interacts with the different roles students assume in feedback activities, particularly the roles of feedback provider and feedback receiver. Since these roles involve distinct cognitive and social demands, it is plausible that specific orientation dimensions may become more salient in one role than in the other. Examining such role-dependent dynamics could provide a deeper understanding of how dispositional factors influence peer-feedback processes. Third, the scale offers opportunities for investigating how peer-feedback orientation develops over time and how it can be influenced through instructional design or pedagogical interventions. Longitudinal studies could examine whether repeated participation in structured peer-feedback activities strengthens particular orientation dimensions, such as feedback self-efficacy or communicative readiness.

Finally, future research should examine the generalizability of the measurement model across different educational contexts, disciplines, and cultural settings. Testing additional levels of measurement invariance across these contexts would provide further evidence for the robustness of the proposed construct and support its application in comparative research on peer-feedback practices.

Acknowledgements This study was supported by a research semester of the first author. Furthermore, this study would have not been possible without the support of several lecturers of the mandatory media education module at Heidelberg University of Education. Special thanks to the lecturers Manuel Altenkirch, Sergej Gil, Katharina Göttmann, Kerstin Liesegang, Dr. Stefan Pietrusky and Julia Werner.

Author Contributions Marco Kalz has contributed to all phases of the research. Annette Schulze has contributed to the conceptualization, resources, project administration and writing of the original draft. Holger Meeh has contributed to the conceptualization, resources, project administration and writing of the original draft. Joshua Weidlich has contributed to the conceptualization, methodology, formal analysis, investigation and writing of the original draft and final manuscript. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. Authors have not received any funding for the study. The work was supported through a research sabbatical of the first author. Open Access funding was enabled and organized by Projekt DEAL.

Data Availability The data and code for the analyis will be made available through the following URI: https://osf.io/zqk8p/.

Declarations

Conflict of interest The authors disclose that they do not have any conflict of interest to declare.

References

Ajjawi, R., & Boud, D. (2017). Researching feedback dialogue: An interactional analysis approach. *Assessment & Evaluation in Higher Education, 42*(2), 252–265. https://doi.org/10.1080/02602938.2015.1102863

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

Aust, F., & Barth, M. (2022). papaja: Prepare reproducible APA journal articles with R Markdown. Retrieved from https://github.com/crsh/papaja, Accessed March 6, 2024

Barth, M. (2022). tinylabels: Lightweight variable labels. Retrieved from https://cran.rproject.org/package=tinylabels, Accessed March 6, 2024

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

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

Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review, 32*(2), 481–509. https://doi.org/10.1007/s10648-019-09510-3

Döring, N. (2013). Zur operationalisierung von geschlecht im fragebogen: Probleme und lösungsansätze aus sicht von mess-, umfrage-, gender-und queer-theorie. *GENDER– Zeitschrift Für Geschlecht, Kultur und Gesellschaft, 5*(2), 17–18.

Friedman, B. A., Cox, P. L., & Maher, L. E. (2008). An expectancy theory motivation approach to peer assessment. *Journal of Management Education, 32*(5), 580–612. https://doi.org/10.1177/1052562907310641

Gielen, S., Dochy, F., & Onghena, P. (2011). An inventory of peer assessment diversity. *Assessment & Evaluation in Higher Education, 36*(2), 137–155. https://doi.org/10.1080/02602930903221444

Greisel, M., Hornstein, J., & Kollar, I. (2025). Do students' beliefs and orientations toward peer feedback predict peer feedback quality and perceptions? *Studies in Educational Evaluation, 84*, Article 101438. https://doi.org/10.1016/j.stueduc.2024.101438

Heidari, S., Babor, T. F., De Castro, P., Tort, S., & Curno, M. (2016). Sex and gender equity in research: Rationale for the SAGER guidelines and recommended use. *Research Integrity and Peer Review, 1*, 1–9. https://doi.org/10.1186/s41073-016-0007-6

Heine, C., & Vesterager, A. (2024). Refining the definition of student peer feedback: The role of roles. *HERMES-Journal of Language and Communication in Business, 64*, 287–300. https://doi.org/10.7146/hjlcb.vi64.153164

Henderson, M., Ryan, T., & Phillips, M. (2019). The challenges of feedback in higher education. *Assessment & Evaluation in Higher Education, 44*(8), 1237–1252. https://doi.org/10.1080/02602938.2019.1599815

Hoaglin, D. C., & Iglewicz, B. (1987). Fine-tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association, 82*(400), 1147–1149. https://doi.org/10.1080/01621459.1987.10478551

Huisman, B., Saab, N., Van Driel, J., & Van Den Broek, P. (2020). A questionnaire to assess students' beliefs about peer feedback. *Innovations in Education and Teaching International, 57*(3), 328–338. https://doi.org/10.1080/14703297.2019.1630294

Jackson, D. L., Gillaspy, J. A., Jr., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods, 14*(1), 6. https://doi.org/10.1037/a0014694

Kane, M. T. (2006). Validation. In R. B. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). American Council on Education/Praeger.

Kasch, J., Van Rosmalen, P., Henderikx, M., & Kalz, M. (2022). The factor structure of the peer-feedback orientation scale (PFOS): toward a measure for assessing students' peer-feedback dispositions. Assessment & Evaluation in Higher Education, 47(1), 15-28. https://doi.org/10.1080/02602938.2021.1893650

Kasch, J., van Rosmalen, P., Kalz, M. (2023). A Thematic Analysis of Factors Influencing Student's Peer-Feedback Orientation. In: Noroozi, O., De Wever, B. (eds) The Power of Peer Learning. Social Interaction in Learning and Development. Springer, Cham. https://doi.org/10.1007/978-3-031-29411-2_12

Kerman, N. T., Banihashem, S. K., Karami, M., Er, E., Van Ginkel, S., & Noroozi, O. (2024). Online peer feedback in higher education: A synthesis of the literature. *Education and Information Technologies, 29*(1), 763–813. https://doi.org/10.1007/s10639-023-12273-8

King, P. E., Schrodt, P., & Weisel, J. J. (2009). The instructional feedback orientation scale: Conceptualizing and validating a new measure for assessing perceptions of instructional feedback. Communication Education, 58(2), 235–261 https://doi.org/10.1080/03634520802515705

Kyriazos, T. A. (2018a). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology, 9*(08), Article 2207. https://doi.org/10.4236/psych.2018.98126

Kyriazos, T. A. (2018b). Applied psychometrics: Writing-up a factor analysis construct validation study with examples. *Psychology, 9*(11), Article 2503. https://doi.org/10.4236/psych.2018.911144

Kyriazos, T., & Poga-Kyriazou, M. (2023). Applied psychometrics: Estimator considerations in commonly encountered conditions in CFA, SEM, and EFA practice. *Psychology, 14*(5), 799–828. https://doi.org/10.4236/psych.2023.145043

Kyriazos, T., & Poga, M. (2023). Dealing with multicollinearity in factor analysis: The problem, detections, and solutions. *Open Journal of Statistics, 13*(3), 404–424. https://doi.org/10.4236/ojs.2023.133020

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*, 936–949. https://doi.org/10.3758/s13428-015-0619-7

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

Liu, N.-F., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279–290. https://doi.org/10.1080/13562510600680582

London, M., & Smither, J. W. (2002). Feedback orientation, feedback culture, and the longitudinal performance management process. *Human Resource Management Review, 12*(1), 81–100. https://doi.org/10.1016/S1053-4822(01)00043-2

Luong, R., & Flake, J. K. (2022). Measurement invariance testing using confirmatory factor analysis and alignment optimization: A tutorial for transparent analysis planning and reporting. *Psychological Methods*. https://doi.org/10.1037/met0000441

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), 741–749. https://doi.org/10.1037/0003-066X.50.9.741

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

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment & Evaluation in Higher Education, 39*(1), 102–122. https://doi.org/10.1080/02602938.2013.795518

Nieminen, J. H., & Carless, D. (2022). Feedback literacy: A critical review of an emerging concept. *Higher Education*. https://doi.org/10.1007/s10734-022-00895-9

Panadero, E. (2016). Is it safe? Social, interpersonal, and human effects of peer assessment: A review and future directions. In G. T. L. Brown & L. R. Harris (Eds.), *Handbook of Social and Human Conditions in Assessment* (pp. 247–266). Routledge.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team. (2023). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/, Accessed March 6, 2024

Reeves, T. D., & Marbach-Ad, G. (2016). Contemporary test validity in theory and practice: A primer for discipline-based education researchers. *CBE—Life Sciences Education, 15*(1), Article rm1. https://doi.org/10.1187/cbe.15-08-0183

Rodríguez, M. F., Nussbaum, M., Yunis, L., Reyes, T., Alvares, D., Joublan, J., & Navarrete, P. (2022). Using scaffolded feedforward and peer feedback to improve problem-based learning in large classes. *Computers & Education, 182*, Article 104446. https://doi.org/10.1016/j.compedu.2022.104446

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21 Word Solution (October 14, 2012). https://doi.org/10.2139/ssrn.2160588

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

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2013). *Using multivariate statistics*. Pearson.

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*, 467–481. https://doi.org/10.1007/s10734-017-0220-3

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249–276. https://doi.org/10.3102/00346543068003249

Topping, K. (2017). Peer assessment: Learning by judging and discussing the work of other learners. *Interdisciplinary Education and Psychology, 1*(1), 1–17. https://doi.org/10.31532/InterdiscipEducPsychol.1.1.007

Topping, K. J. (2009). Peer assessment. *Theory into Practice, 48*(1), 20–27. https://doi.org/10.1080/00405840802577569

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

UNESCO. (2023). *The teachers we need for the education we want. The global imperative to reverse the teacher shortage*. United Nations Educational, Scientific and Cultural Organization.

Wen, M. L., & Tsai, C.-C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education, 51*, 27–44. https://doi.org/10.1007/s10734-004-6375-8

Wild, E., & Esdar, W. (2014). Eine heterogenitätsorientierte Lehr-/Lernkultur für eine Hochschule der Zukunft. Fachgutachten im Auftrag des Projektes Nexus der Hochschulrektorenkonferenz.

Woitt, S., Weidlich, J., Jivet, I., Orhan Göksün, D., Drachsler, H., & Kalz, M. (2025). Students' feedback literacy in higher education: an initial scale validation study. Teaching in Higher Education, 30(1), 257-276. https://doi.org/10.1080/13562517.2023.2263838

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

Zhu, Q., & Carless, D. (2018). Dialogue within peer feedback processes: Clarification and negotiation of meaning. *Higher Education Research & Development, 37*(4), 883–897. https://doi.org/10.1080/07294360.2018.1446417

Authors and Affiliations

Marco Kalz[1] · Annette Schulze[1] · Holger Meeh[1] · Joshua Weidlich[2,3]

✉ Marco Kalz
kalz@ph-heidelberg.de

[1] Department of Educational Technology, Institute for Arts, Music and Media, Faculty of Cultural Science and Humanities, Heidelberg University of Education, Im Neuenheimer Feld 561, 69120 Heidelberg, Germany

[2] University of Zurich, Zurich, Switzerland

[3] Zurich University of Teacher Education, Zurich, Switzerland