

Title (max 21 words)

The Measurement of Student Engagement: Qualitative Validity Evidence for a Novel German-language Scale

Proposal Information, e.g., topic, research question, objective, conceptual and theoretical framework (600 words max)

With a sixfold increase in the number of publications between 2010 and 2020 alone (Salmela-Aro et al., 2021), the construct of student engagement continues to attract both European and international research interest. Defined as a meta-construct encompassing students' behavioural, cognitive and emotional reactions to school (Fredricks et al., 2004), it is widely regarded as a robust predictor of academic performance and an important mediator of various contextual factors that influence children's adjustment to school (e.g., Reschly & Christenson, 2022; Skinner & Raine, 2022; Wong et al., 2024). The currently existing body of research on engagement in the German school-context is limited, and the measures used in these studies have typically been translated from international literature (e.g., Bakadorova et al., 2020; Hagenauer et al., 2015; Hoferichter & Raufelder, 2023; Preusche & Göbel, 2022; Raufelder et al., 2013; Teuber et al., 2020). Therefore, a student engagement scale that is freely available, widely applicable, and validated for use with German students has particular potential to stimulate new research interest.

A German-language self-report scale measuring student engagement from late primary to early secondary school was developed and validated using a mixed-methods approach (McCullagh, 2026). Confirmatory factor analyses ($N = 558$) indicate that the scale consists of two correlated factors (affective engagement and cognitive-behavioral engagement), with an uncorrelated method-bifactor accounting for the impact of item valence. Partial measurement invariance was confirmed across gender, age, home language (German, other), and school track (Gymnasium/high-track, other). Supplementing these general quantitative findings, the presentation will concentrate on the qualitative evidence to further support the construct validity of the new measure. Results from cognitive interviews conducted with a sample of 15 students in grades 4 through 7 are presented. An evaluative content analysis was conducted (Kuckartz & Rädiker, 2022), pursuing two interlocking objectives. The initial objective was to ascertain whether the selected engagement indicators effectively measure the intended behaviors and emotions. For this purpose, the analysis drew upon an intricate coding framework ($\kappa = .65$) to assess respondents' accuracy with respect to the comprehension of item content. Secondly, the individual applicability of the quantitative measurement model was investigated by incorporating response consistency, defined as the tendency of interviewees to respond similarly to items intended to measure related content. Identification of instances in which students deviate from this expected pattern allows for a unique examination of the circumstances under which the current quantitative model may have limited applicability. Narrative and case-based summaries are employed to formulate hypotheses for the adjustment of theoretical and measurement models in future quantitative research.

Method/Instruments (400 max)

Semi-structured interviews were conducted in December of 2024 with an initial cohort of 11 students (grades 4-6) recruited from a primary and a comprehensive school in North-Rhine Westphalia, Germany. To enhance the heterogeneity of the sample, four additional interviews with a second cohort of students encountering learning difficulties were carried out in February 2025. Participants were presented with multiple sets of engagement indicators ($N = 24$). Students were encouraged to think "out loud", i.e., articulate their reflections on the content and language of each item, as well as the rationale underlying their quantitative rating. The interviewer provided prompts and follow-up questions to encourage further dialogue. Interviews were audio-recorded and ranged in duration from

11 to 34 minutes. An evaluative content analysis of interview transcripts was conducted using MAXQDA. Two criteria were employed: response accuracy and response consistency.

Accuracy reflected the extent of congruence between students' comprehension of an engagement indicator and conceptual definitions, delineated in the coding framework. Coders evaluated the accuracy of students' responses to each individual item using a four-point Likert scale (fully on-target – fully off-target; $\kappa = 0.65$, discrepancies between coders resolved by assigning the lower rating). For instance, a student's response to the item "I think deeply about what we are learning in class" would be coded as fully on-target if it referred to (or provided an example of) actively considering the solution/requirements of a task, or purposefully exerting mental effort to comprehend information. Statements that were only tangentially related to (or irreconcilable with) this general state of affairs were considered off-target.

The premise underlying the evaluation of response consistency was that, given applicability of the measurement model, participants would respond in a comparable manner to indicators that were closely related in content. By contrast, an inconsistent response was defined as a difference of more than one scale point within a shared content dimension (e.g., within items capturing "participation") or more than two scale points between content dimensions (e.g., between "participation" and "attention"). For instance, if a student indicated that they "almost always" paid attention in class, it would be flagged as an unusual response pattern if they simultaneously indicated that they are "sometimes" distracted (>1 difference within-category) or "seldom" participated in class (>2 difference between categories). Narrative and case-based summaries were conducted to identify common themes among responses that diverged from theoretical expectations with regard to accuracy, consistency or both.

Conclusions/Findings (Max 300 words)

Response accuracy was high across the first (88.52%) and second cohort (84.21%). The most common reason for inaccurate responses was the application of theoretically inappropriate frames of reference (e.g., equating engagement with academic achievement). However, these errors were randomly distributed across items. Therefore, they are unlikely to be attributable to flaws in the operationalization of engagement indicators.

Response consistency was substantially lower in the second cohort (73.6%) compared to the first (97.15%). Specifically, the second cohort of respondents seemed to both under- and overestimate their engagement at various points. Furthermore, the rationales provided by students to explain these atypical rating patterns frequently exhibited contradictions. Case-based summaries revealed that students in the second cohort conditioned their engagement on a variety of internal and external factors, including mood, the characteristics of the task, and pedagogical approaches. Arguably, the engagement of this cohort is more accurately characterized as a fluctuating state rather than a stable trait. Preliminary findings further suggest that the theoretical dimensions of engagement may exhibit greater variability in this cohort. For instance, effort and attention manifested as partially independent variables for some, despite both being considered indicators of behavioral engagement.

In general, the current engagement measure demonstrates high construct validity, such that even the youngest students are generally able to provide meaningful responses that are in line with theoretical expectations. Nevertheless, case-based analysis indicates that trait-like self-report measures of engagement, such as the current scale, may demonstrate a systematic decrease in reliability among students encountering academic difficulties. Research focusing on such cohorts should recruit larger samples, closely examine measurement invariance, and explore alternative measurement models where appropriate. If these limitations are kept in mind, the current scale will function as a reliable and valid measure of student engagement for a range of research interests.

References

- Bakadorova, O., Lazarides, R., & Raufelder, D. (2020). Effects of social and individual school self-concepts on school engagement during adolescence. *European Journal of Psychology of Education*, 35(1), 73–91. <https://doi.org/10.1007/s10212-019-00423-x>
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>
- Hagenauer, G., Hascher, T., & Volet, S. E. (2015). Teacher emotions in the classroom: associations with students' engagement, classroom discipline and the interpersonal teacher-student relationship. *European Journal of Psychology of Education*, 30(4), 385–403. <https://doi.org/10.1007/s10212-015-0250-0>
- Hoferichter, F., & Raufelder, D. (2023). Biophysiological stress markers relate differently to grit and school engagement among lower- and higher-track secondary school students. *The British Journal of Educational Psychology*, 93(S1), 174–194. <https://doi.org/10.1111/bjep.12514>
- Kuckartz, U., & Rädiker, S. (2022). *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung: Grundlagentexte Methoden* (5. Auflage). *Grundlagentexte Methoden*. Beltz Juventa. http://www.content-select.com/index.php?id=bib_view&ean=9783779955337
- McCullagh, L. (2026). *The Measurement of Student Engagement: A Critical Examination of Operational Definitions and the Development of a German-Language Scale* [unpublished Doctoral dissertation, Universität Münster].
- Preusche, Z. M., & Göbel, K. (2022). Does a Strong Bicultural Identity Matter for Emotional, Cognitive, and Behavioral Engagement? *Education Sciences*, 12(1), 5. <https://doi.org/10.3390/educsci12010005>
- Raufelder, D., Kittler, F., Braun, S. R., Lätsch, A., Wilkinson, R. P., & Hoferichter, F. (2013). The interplay of perceived stress, self-determination and school engagement in adolescence. *School Psychology International*, 35(4), 405–420. <https://doi.org/10.1177/0143034313498953>
- Reschly, A. L., & Christenson, S. L. (2022). Epilogue. In A. L. Reschly & S. L. Christenson (Eds.), *Handbook of Research on Student Engagement* (pp. 659–666). Springer International Publishing.
- Salmela-Aro, K., Tang, X., Symonds, J., & Upadaya, K. (2021). Student Engagement in Adolescence: A Scoping Review of Longitudinal Studies 2010-2020. *Journal of Research on Adolescence : The Official Journal of the Society for Research on Adolescence*, 31(2), 256–272. <https://doi.org/10.1111/jora.12619>
- Skinner, E. A., & Raine, K. E. (2022). Unlocking the Positive Synergy Between Engagement and Motivation. In A. L. Reschly & S. L. Christenson (Eds.), *Handbook of Research on Student Engagement* (pp. 25–56). Springer International Publishing. https://doi.org/10.1007/978-3-031-07853-8_2
- Teuber, Z., Möer, J., Webel, L., & Seifert, A. (2020). Burnout und Engagement: Eine Pilotstudie zum Transfer des Job-Demands-Resources-Modells in den deutschen Schulkontext. *Prävention Und Gesundheitsförderung*, 15(3), 269–274. <https://doi.org/10.1007/s11553-019-00752-1>
- Wong, Z. Y., Liem, G. A. D., Chan, M., & Datu, J. A. D. (2024). Student engagement and its association with academic achievement and subjective well-being: A systematic review and meta-analysis. *Journal of Educational Psychology*, 116(1), 48–75. <https://doi.org/10.1037/edu0000833>