ASSESSING MATHEMATICAL MODELLING COMPETENCIES WITH DIGITAL PORTFOLIOS

Catharina Beckschulte, Jascha Quarder, Annika Rosendahl, Stanislaw Schukajlow

University of Münster

Digital portfolios have been proposed as a promising assessment format for mathematical modelling. However, empirical evidence on whether such portfolios can be used to reliably and validly assess analytically defined modelling sub-competencies is limited. We investigated the development and validation of an assessment instrument that combines digital portfolios based on holistic modelling tasks with an analytic assessment rubric. Using data from 116 ninth-grade students, we examined the internal structure, interrater reliability, and criterion-related validity. Results indicate that modelling sub-competencies can be empirically distinguished, assessed with good reliability, and meaningfully related to an established modelling test, suggesting that digital portfolios can bridge holistic task design and analytic assessment.

INTRODUCTION

Mathematical modelling is widely regarded as a key competency in mathematics education and is explicitly anchored in international curricula as a competency that should be both taught and assessed (Niss & Blum, 2020). Accordingly, various assessment approaches have been developed, ranging from holistic measures of overall modelling competence to analytic instruments targeting specific modelling sub-competencies (Cevikbas et al., 2022). Standardized written tests have proven useful for assessing analytic sub-competencies, such as simplifying, mathematizing, interpreting, and validating (Hankeln et al., 2019). However, it is challenging for such written test formats to adequately represent the openness and authenticity of holistic modelling tasks, which are commonly characterized by these features (Frejd, 2013; Krawitz et al., 2025). At the same time, such holistic tasks require analytic assessment rubrics to make different facets of students' modelling competencies systematically visible (Borromeo Ferri & Greefrath, 2026). Against this background, (digital) portfolios have been discussed as an alternative assessment format that may support the documentation and evaluation of students' modelling processes, particularly when digital tools are involved (Frejd, 2013; Yang et al., 2017). However, empirical evidence on whether digital portfolio-based assessments that are based on holistic modelling tasks can be used to reliably and validly measure analytically defined modelling sub-competencies remains scarce. Thus, we investigated whether an analytic assessment rubric applied to a digital portfolio is suitable for measuring students' modelling competencies at the level of sub-competencies, suggesting that digital portfolios may serve as a bridge between holistic modelling tasks and analytic assessment approaches.

2026. In H. Krzywacki, M. Hähkiöniemi, E. Haataja, M.S. Hannula & P. Hästö (Eds.). *Proceedings of the 49th Conference of the International Group for the Psychology of Mathematics Education: Research Reports*, pp. 155-162. PME.

THEORETICAL BACKGROUND

Mathematical Modelling Competencies

Research on mathematical modelling competencies is commonly characterized by two approaches, a top-down (holistic) and a bottom-up (analytic) perspective, which differ in how modelling competence is defined and operationalized (Cevikbas et al., 2022). The top-down approach conceptualizes modelling competence as a single holistic construct, whereas the bottom-up approach conceptualizes it as a composite of sub-competencies linked to phases of the modelling cycle (Cevikbas et al., 2022). Drawing on psychological definitions of competence, this perspective distinguishes between global modelling competence and analytically defined sub-competencies (e.g., simplifying, mathematizing, working mathematically, interpreting, validating) (Blum & Leiß, 2007), which are necessary for carrying out and reflecting on modelling processes. At the same time, mathematical modelling competence is also understood to involve metacognitive, creative, motivational, and social dimensions, particularly when students engage with complex and open modelling tasks (Lu et al., 2025; Stillman, 2011). In the bottom-up perspective, modelling is characterized via sub-competencies that can be operationalized, empirically distinguished, and assessed with analytic instruments (Hankeln et al., 2019). This analytic view has been adopted in quantitative research aiming to investigate learning processes, instructional effects, or assessment formats related to mathematical modelling (Cevikbas et al., 2022).

Assessing Mathematical Modelling Competencies with Tests and Portfolios

Depending on whether modelling competence is understood as a holistic ability or as a set of analytically distinguishable sub-competencies, different assessment formats and measurement strategies are employed (Cevikbas et al., 2022; Frejd, 2013). Established assessment instruments often rely on written tests that target modelling sub-competencies that are aligned with modelling cycle phases. Such tests have been used successfully to assess abilities such as simplifying, mathematizing, interpreting, and validating (Hankeln et al., 2019). However, written test formats typically decompose modelling tasks into predefined steps and focus on short responses or final results, thus limiting their capacity to capture the non-linear and open-ended nature of authentic modelling processes (Frejd, 2013). Furthermore, with this approach, it is challenging to ensure that a task targets exactly the intended sub-competency and no other sub-competencies, thereby posing a threat to construct validity (Hankeln et al., 2019). While holistic approaches are well suited to represent complete modelling processes, they pose challenges to reliability, comparability, and feasibility, whereas analytic approaches enable more precise diagnostic insights but risk fragmenting modelling competence (Cevikbas et al., 2022). To assess modelling competencies, several authors have highlighted the importance of assessment rubrics that are explicitly aligned with phases of the modelling process. Such rubrics aim to support systematic evaluation of open-ended modelling solutions by making assessment

criteria explicit and by guiding judgments across different modelling phases (Borromeo Ferri & Greefrath, 2026).

Beyond assessment criteria and scoring procedures, assessment formats differ in the extents to which they can be used to document and evaluate students' modelling processes. Portfolios represent a process-oriented assessment format and are commonly defined as structured collections of student-generated artefacts documenting learning processes and outcomes (Frejd, 2013). Digital portfolios, in particular, support multimodal documentation and the integration of digital tools (Yang et al., 2017), which are difficult to capture in paper-and-pencil test formats. However, portfolio-based assessments pose challenges for reliability and validity of assessment inferences, as the open-ended nature of student-generated artefacts and reliance on teacher judgment may lead to variability in scoring and limited comparability (Krall, 2023). Empirical evidence on the measurement properties of portfolio-based assessments in mathematical modelling therefore remains scarce, especially with regard to assessing sub-competencies with analytical assessment rubrics.

RESEARCH QUESTIONS AND EXPECTATIONS

This study was part of the AlPMo project (Alternative Forms of Assessment in Mathematical Modelling, in German: Alternative Prüfungsformen beim mathematischen Modellieren (Beckschulte & Quarder, 2025)). In this project, a digital portfolio is developed and evaluated as an assessment format for mathematical modelling in classroom contexts. The portfolio is designed to document students' work on holistic modelling tasks and serve as a basis for assessing modelling competencies with an analytic assessment rubric. Building on bottom-up conceptualizations of modelling competence and prior research on the assessment of modelling sub-competencies, we examined whether this portfolio-based assessment allows for the reliable and valid measurement of modelling sub-competencies. To address this aim, the following research questions were formulated:

RQ1: To what extent do the theoretically assumed sub-competencies of mathematical modelling emerge as empirically distinguishable dimensions in students' digital portfolios developed in response to a holistic modelling task?

RQ2: To what extent can students' mathematical modelling performance in digital portfolios be assessed reliably across these empirically identified dimensions?

RQ3: To what extent is students' mathematical modelling performance in digital portfolios related to their performance on an established modelling test?

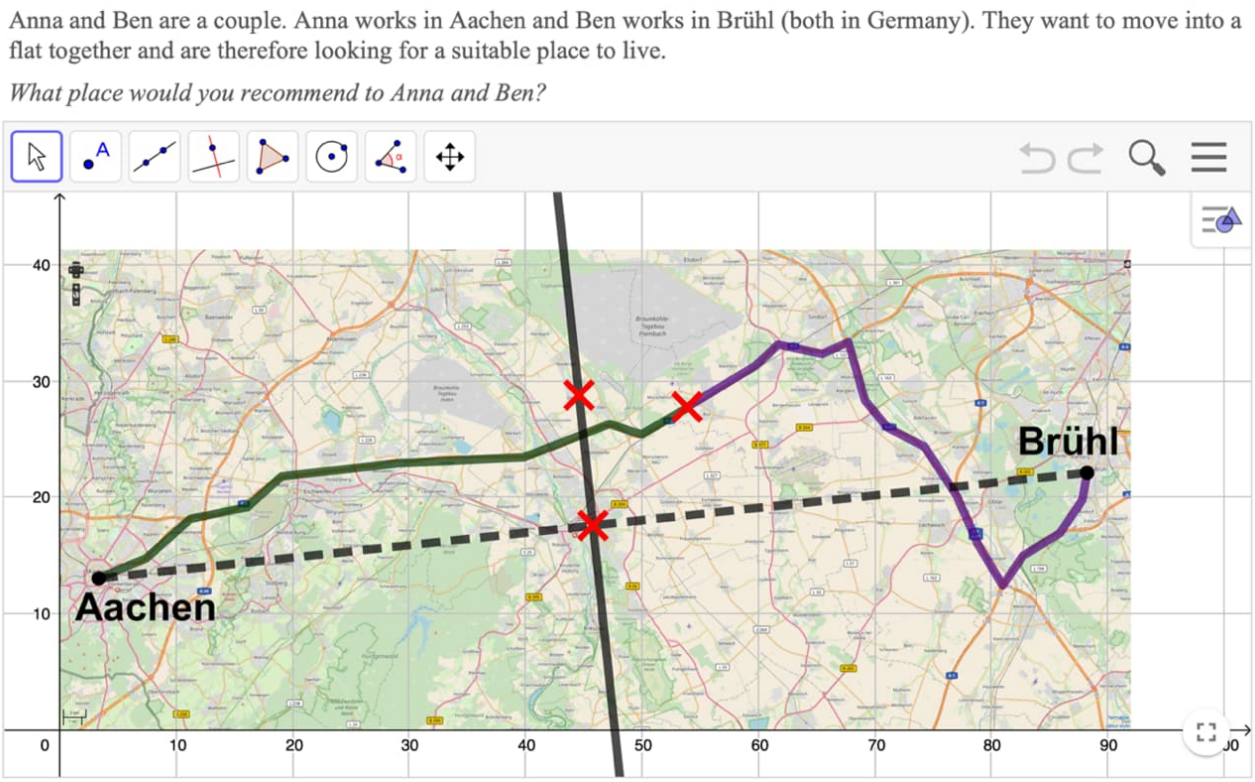
On the basis of theoretical assumptions about analytically defined sub-competencies (Blum & Leiß, 2007; Cevikbas et al., 2022), we expected that students' portfolio performance would be better represented by a multidimensional structure than by a unidimensional one. Also, drawing on prior empirical research on the assessment of modelling sub-competencies and analytic scoring approaches (Borromeo Ferri & Greefrath, 2026; Hankeln et al., 2019), we expected that the use of an analytic

assessment rubric would result in satisfactory interrater reliability and positive relationships between portfolio-based sub-competency measures and external criteria.

METHOD

The study was conducted with $N = 116$ ninth-grade students attending a German secondary school (Gymnasium) at the lower secondary level (Grade 9). Data were collected in the context of regular mathematics lessons in which a digital portfolio as an alternative assessment format was implemented.

Figure 1. The Flat Search task with three possible solutions (Beckschulte & Quarder, 2025). Screenshot © GeoGebra.org; map data © OpenStreetMap contributors.



Each student created an individual digital portfolio documenting their work on two mathematical modelling tasks over the course of a five-lesson teaching unit. The portfolios were created with a note-taking application on a tablet and included written explanations, mathematical calculations, and screenshots from GeoGebra applets used in the modelling process. At the end of the unit, the portfolios were submitted as PDF files and served as the basis for assessment with the assessment rubric. The modelling tasks embedded in the portfolio were open, real-world problems designed to elicit holistic modelling processes. For this study, two tasks were used: one focusing on geometric modelling in the context of estimating the volume of a real-world pool, and one addressing spatial modelling in a location-based decision-making context, the Flat Search task (see Figure 1). Both tasks required students to work through the central phases of mathematical modelling, including simplifying the real-world situation,

developing and applying a mathematical model, and checking and reflecting on the obtained solution. Digital tools were an integral part of task processing, as students used GeoGebra to construct models, perform calculations, and explore the effects of varying assumptions.

Students' portfolio work was assessed with a rubric that was based on established frameworks for assessing mathematical modelling competencies and adapted to the characteristics of digital portfolios. The rubric comprised four dimensions: (1) Written presentation (*P_Pres*), (2) Construction of a model (*P_Model*), (3) Working mathematically (*P_Math*), and (4) Validating (*P_Val*). *P_Pres* addressed the quality of the portfolio's written work, focusing on structural coherence, linguistic accuracy, clarity, and the appropriate use of representations, and was therefore defined as a portfolio-specific assessment dimension rather than a modelling sub-competency. *P_Model* focused on students' ability to simplify the given real-world situation, make appropriate assumptions, and mathematize the real model by selecting suitable mathematical approaches. *P_Math* addressed the correctness of mathematical procedures, calculations, and the use of digital tools. *P_Val* captured students' reflection on the plausibility of their results, their critical evaluation of assumptions, and their consideration of alternative modelling approaches. Each dimension was operationalized through several task-specific criteria, allowing for an analytic assessment of students' modelling performance across different facets.

The analysis followed a three-step approach aligned with the three research questions: Construct validity (RQ1) was examined by analyzing the internal structure of the portfolio-based assessment using confirmatory factor analysis with the WLSMV estimator (Brown, 2015), comparing a theoretically specified four-dimensional model (*P_Pres*, *P_Model*, *P_Math*, *P_Val*) with a unidimensional model. Model comparisons were conducted with chi-square difference tests appropriate for ordinal data. In addition, overall model fit was evaluated with established fit indices (Wilson, 2008). Interrater reliability (RQ2) was assessed by calculating intraclass correlation coefficients (ICCs) for each rubric dimension based on a double-coded subsample of 36 portfolios. This subsample was independently coded by two trained raters. Criterion-related validity (RQ3) was examined by relating portfolio-based measures to the sub-competencies of simplifying (*M_Simp*), mathematizing (*M_Math*), interpreting (*M_Int*), and validating (*M_Val*) as assessed with an established modelling test (Hankeln et al., 2019). In line with theoretical assumptions, associations were analyzed between the corresponding portfolio and the sub-competencies on the test.

RESULTS

As expected, the model comparison for examining the construct validity of the portfolio-based assessment (RQ1) indicated that the four-dimensional model provided a better fit to the data than the unidimensional model. This was supported by a significant result of the robust chi-square difference test ($\Delta\chi^2 = 49.36$, $\Delta df = 6$, $p < .001$). Model fit indices for the four-dimensional model indicated a good overall fit

(CFI = .98, TLI = .97, RMSEA = .06). Latent correlations between the content-related dimensions were high (r = .54–.86) but remained within an acceptable range.

To test interrater reliability (RQ2), we examined ICCs between the coding for each rubric dimension based on the ratings of two independent coders. The following ICC values were obtained: *P_Pres* (ICC = .81), *P_Model* (ICC = .86), *P_Math* (ICC = .84), and *P_Val* (ICC = .82). According to commonly used guidelines (Koo & Li, 2016), these values indicate good interrater reliability across all dimensions.

Table 1: Correlations between portfolio dimensions and modelling sub-competencies

Dimension	M_Simp	M_Math	M_Int	M_Val
P_Pres	.01	.17*	-.18*	.00
P_Model	.03	.20*	-.16*	-.02
P_Math	-.01	.25**	-.13	.02
P_Val	-.01	.20*	-.15	.11

Note. *p < .05, **p < .01, ***p < .001; one-tailed tests.

Criterion-related validity of the portfolio-based assessment (RQ3) was examined by relating portfolio performance to an established modelling test for assessing modelling sub-competencies (Hankeln et al., 2019). Table 1 summarizes the correlations between the four portfolio dimensions and the corresponding sub-competencies assessed with the test. Based on directed hypotheses, one-tailed tests were used. Significant positive correlations were found between all portfolio dimensions and mathematizing, while significant negative correlations emerged for interpreting, and no significant associations were observed for simplifying or validating.

DISCUSSION AND IMPLICATIONS

This study examined whether an assessment instrument combining digital portfolios with an analytic assessment rubric can be used to reliably and validly measure mathematical modelling sub-competencies. Across all three research questions, the results provide converging evidence that portfolio-based assessment instruments can support a differentiated measurement of modelling competencies. First, the analyses of the internal structure indicated that students' portfolio performances were better represented by a multidimensional model than by a unidimensional model, supporting the theoretical distinction between the dimensions Written presentation, Construction of a model, Working mathematically, and Validating. Second, interrater reliability estimates reached satisfactory levels across rubric dimensions, suggesting that structured scoring can yield reliable judgments even for open and process-oriented student artefacts. Third, meaningful relationships between portfolio-based measures

and an established modelling test were found, indicating that digital portfolios capture theoretically related but not identical aspects of students' modelling performance.

From a theoretical perspective, the emergence of empirically distinguishable sub-competencies in students' digital portfolios contributes to ongoing discussions on the assessment of mathematical modelling competencies, indicating that holistic modelling tasks can provide a suitable basis for analytic assessment approaches. In this sense, the results align with bottom-up conceptualizations of modelling competence that assume analytically separable sub-competencies related to phases of the modelling cycle (Blum & Leiß, 2007; Cevikbas et al., 2022). The findings further suggest that digital portfolios may bridge holistic task design and analytic measurement by enabling open modelling work alongside structured, theory-based scoring through an analytic assessment rubric (Borromeo Ferri & Greefrath, 2026). By showing empirically distinguishable dimensions and meaningful relationships with an established modelling test (Hankeln et al., 2019), this study extends prior validation work based on standardized tests to more process-oriented, digitally supported assessment settings.

With regard to assessment research, the findings suggest that digital portfolios allow for a differentiated assessment of students' modelling sub-competencies while preserving the openness and authenticity of holistic modelling tasks (Krawitz et al., 2025). Portfolio-based assessments may complement established test formats by capturing students' documented modelling work and reflections across different phases of the modelling process beyond what is typically assessed with standardized written tests (Frejd, 2013). Thus, for classroom practice, portfolios may be a feasible option for assessing modelling competencies in digitally enriched learning environments. However, the results need to be interpreted in light of limitations, including the focus on a single school context, a limited number of modelling tasks, and a specific assessment rubric. Future research should therefore examine the robustness of the proposed assessment approach across different modelling contexts, task types, and student populations and complement quantitative validation approaches with qualitative analyses to gain deeper insights into which aspects of modelling competence are captured in students' digital portfolios.

References

Beckschulte, C., & Quarder, J. (2025). *Alternative forms of assessment in mathematical modelling: The AlPMo project in practice and research.* University of Münster (MIAMI – Institutional Repository). https://doi.org/10.17879/02958629587

Blum, W., & Leiß, D. (2007). How do students and teachers deal with modelling problems? In C. Haines, P. Galbraith, W. Blum, & S. Khan (Eds.), *Mathematical modelling* (pp. 222–231). Woodhead Publishing. https://doi.org/10.1533/9780857099419.5.221

Borromeo Ferri, R., & Greefrath, G. (2026). Pre-service teachers' assessment competence for grading students' solutions of mathematical modelling problems. *Teaching and Teacher Education, 169*, 105281. https://doi.org/10.1016/j.tate.2025.105281

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.

Cevikbas, M., Kaiser, G., & Schukajlow, S. (2022). A systematic literature review of the current discussion on mathematical modelling competencies: State-of-the-art developments in conceptualizing, measuring, and fostering. *Educational Studies in Mathematics, 109(2)*, 205–236. https://doi.org/10.1007/s10649-021-10104-6

Frejd, P. (2013). Modes of modelling assessment—A literature review. *Educational Studies in Mathematics, 84(3)*, 413–438. https://doi.org/10.1007/s10649-013-9491-5

Hankeln, C., Adamek, C., & Greefrath, G. (2019). Assessing sub-competencies of mathematical modelling—Development of a new test instrument. In G. Stillman & J. Brown (Eds.), *Lines of inquiry in mathematical modelling research in education* (pp. 143–160). Springer. https://doi.org/10.1007/978-3-030-14931-4_8

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15(2)*, 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Krall, G. (2023). Teacher-designed mathematical portfolio assessments: Motivations, potential benefits, and lessons learned. *Canadian Journal of Science, Mathematics and Technology Education, 23(2)*, 303–321. https://doi.org/10.1007/s42330-023-00280-3

Krawitz, J., Schukajlow, S., Yang, X., & Geiger, V. (2025). A systematic review of international perspectives on mathematical modelling: Modelling goals and task characteristics. *ZDM – Mathematics Education, 57*, 193–212. https://doi.org/10.1007/s11858-025-01683-2

Lu, X., Kaiser, G., Zhu, Y., Ma, H., & Yan, Y. (2025). Mathematical creativity in modelling: Further development of the construct, its measurement, and its empirical implementation. *ZDM – Mathematics Education, 57(2)*, 365–379. https://doi.org/10.1007/s11858-025-01652-9

Niss, M., & Blum, W. (2020). *The learning and teaching of mathematical modelling*. Routledge. https://doi.org/10.4324/9781315189314

Stillman, G. (2011). Applying metacognitive knowledge and strategies in applications and modelling tasks at secondary school. In G. Kaiser, W. Blum, R. Borromeo Ferri, & G. Stillman (Eds.), *Trends in teaching and learning of mathematical modelling* (pp. 165–180). Springer. https://doi.org/10.1007/978-94-007-0910-2_18

Wilson, M. (2008). *Constructing measures—An item response modeling approach*. Routledge. https://doi.org/10.4324/9781410611697

Yang, M., Wang, T., & Lim, C. P. (2017). E-portfolios as digital assessment tools in higher education. In M. J. Spector, B. B. Lockee, & M. D. Childress (Eds.), *Learning, design, and technology: An international compendium of theory, research, practice, and policy* (pp. 1–23). Springer. https://doi.org/10.1007/978-3-319-17727-4_83-1