

# Modelling Competence(s) in Written Comparison Tests

Uwe Schürmann, Georg Bruckmaier

*University of Applied Sciences and Arts Northwestern Switzerland, School of Education*

**Abstract** The present study investigates the extent to which tasks in written comparison tests for Year 8 students in Germany (VERA-8) meet established quality criteria for modelling tasks. Furthermore, it examines whether these tasks are suitable for assessing modelling competence in a theory-based manner, drawing on atomistic and holistic approaches to conceptualising modelling competence. The findings indicate that VERA-8 tasks do not fully meet the established quality criteria for modelling tasks, such as authenticity and openness. Nevertheless, the assessment of modelling competence based on atomistic and holistic approaches is, in principle, a feasible undertaking.

**Keywords** Mathematical modelling, Competence, Assessment, Comparison tests, Atomistic and holistic approaches, Educational standards

## 1 Introduction

Mathematical modelling—the process of solving realistic problems with mathematical means, a process which can be characterised by its interplay between reality and mathematics (Niss et al., 2007)—is a central goal of mathematics education in many countries (Kaiser, 2020). However, analyses of test items from national course tests (Frejd, 2011, 2013) and central examinations (Greefrath et al., 2017; Siller & Greefrath, 2020) indicate that written tests frequently neglect key aspects of modelling. Therefore, even when modelling is firmly embedded in curricula, this does not necessarily guarantee that competences are adequately assessed in examinations. This discrepancy between curricular expectations and assessment practice raises the question of whether it also occurs in other contexts.

Research has examined how modelling can be fostered (Cevikbas et al., 2022) and assessed in different formats (Kaiser, 2007) while written tests remain the primary focus (Frejd, 2013). Beyond such explicitly designed tasks, modelling is also assessed in large-scale assessments (e.g., OECD, 2023), central examinations, and nationwide comparison tests like VERA-8 (German: **V**ergleichs-**A**rbeiten in Jahrgangsstufe **8**, i.e., comparison tests in Year 8). However, it is important to note that these written tests must meet additional educational policy and practical requirements (Drücke-Noe, 2012), which may complicate the systematic assessment of modelling competence.

In this context, the present study examines how modelling competence is assessed in VERA-8, since these tests are designed to ensure that the educational standards defined by the German Standing Conference of the Ministers of Education and Cultural Affairs (KMK, 2004) are implemented in schools, including those related to mathematical modelling. The study is based on the premise that modelling competence is best assessed through tasks that meet established quality criteria, such as authenticity or openness of the task (Maaß, 2010; also see Table 4.1).

## 2 Theoretical Background

Competence is understood as a domain-specific, context-dependent disposition to perform, which can be learnt in principle (Weinert, 2001). Competence diagnostics therefore require an alignment between the internal structure of a competence and the procedures used to assess it. In addressing this question, the present study will consider both the modelling of competence and its subsequent assessment.

### 2.1 Modelling Competence(s)

Competence is essentially defined by knowledge and skills, though debate remains over whether affective aspects should also be included. Thus, modelling competence is sometimes defined *with* (e.g., Kaiser, 2007; Maaß, 2006) or *without* (e.g., Niss & Højgaard, 2019) affective and motivational components. A theoretical reconstruction of modelling competence must be distinguished from individuals' actual performance dispositions, as empirical studies demonstrate its close connection to other competences, for instance problem solving (Greefrath, 2015) and reading (Krawitz et al., 2022).

Definitions of modelling competence can generally be categorised into *holistic* (top-down) and *atomistic* (bottom-up) approaches (Cevikbas et al., 2022). A holistic approach to modelling competence involves the assessment of competence throughout the entire modelling cycle. This approach typically characterises modelling competence in terms of levels. For instance, Greer and Verschaffel's (2007) approach involves the evaluation of the role and utilisation of mathematical modelling across various disciplines and within society. Henning and Keune's (2007) work offers a similar perspective, characterising the highest level as the analysis and reflection on model construction, its purpose, and the criteria for evaluation.

In contrast, atomistic approaches delineate modelling competence through sub-competences corresponding to phases of a diagnostic modelling cycle (e.g., Blum & Leiß, 2007). These typically include understanding real-world problems, developing a real model, constructing and solving a mathematical model, interpreting results, and validating solutions. In accordance with prevailing convention, we refer to modelling *competence* in the singular for the overall process and to modelling *competences* in the plural for its sub-components. In the context of test and task design, the distinction between the two approaches is of practical relevance. Atomistic approaches enable finer-grained diagnostic differentiation, whereas holistic approaches capture integrated performance.

### 2.2 Assessment of Modelling Competence(s)

Here, assessment is understood as the empirical evaluation of individuals' modelling competence(s), excluding purely theoretical arguments, descriptions of classroom practice or grading. Approaches range from subtasks in standardised tests to complex projects, generating quantitative (e.g., multiple-choice responses) or qualitative data (e.g., videos, transcripts, portfolios). These may be static or process-oriented; the latter category includes eye-tracking data (Schindler et al., 2025) and log files from computer-based assessments that

capture click patterns, response sequences, and processing times (Hankeln et al., 2025). In accordance with the principles of competence diagnostics, assessments can be formative to support learning or summative to evaluate performance (see Table 2.1).

Table 2.1: Assessment of modelling competence(s)

Aspect	Characteristic and examples
Range	Atomistic (e.g., sub-tasks) ↔ Holistic (e.g., modelling projects)
Type of empirical data	Quantitative (e.g., Multiple choice resp.) ↔ Qualitative (e.g., videos) Static (e.g., single responses) ↔ Dynamic (e.g., eye-tracking)
Diagnostic function	Formative ↔ Summative

The focus of written tests can be exclusively on modelling (e.g., Hankeln et al., 2019) or assessing it alongside other mathematical competences. Furthermore, such tests vary in their mode of administration, ranging from paper-and-pencil to fully computer-based and hybrid formats. Additionally, there is variation in task types, such as the use of tactile tools (e.g., compasses) in paper-based tests or animations and interactive graphics in computer-based formats (e.g., Wirth & Greefrath, 2024).

### 3 Research questions

In consideration of the aforementioned factors relating to modelling competence(s) and their assessment—and based on the premise that modelling competence(s) are best assessed when students engage in tasks that meet defined criteria—the objective of this study is to determine whether and to what extent modelling skills are assessed by VERA-8 tasks. Furthermore, the present study investigates the additional potential of these tasks for a theory-based assessment of modelling competence(s). Therefore, the present study investigates the following two research questions (RQ):

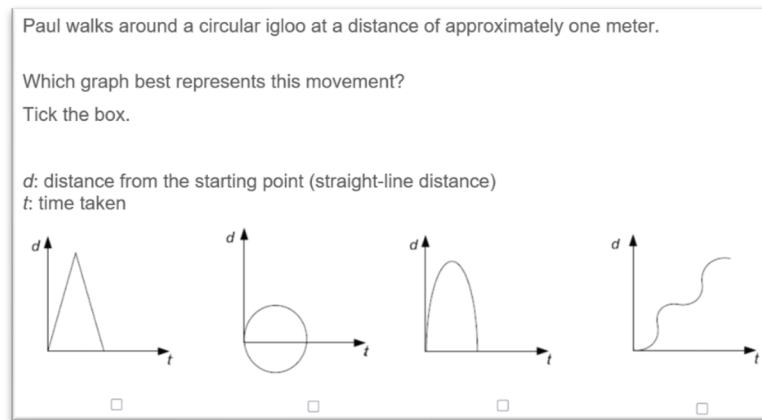
- RQ 1: To what extent do VERA-8 tasks designed to assess modelling competence meet established quality criteria for modelling tasks?
- RQ 2: To what extent do the VERA-8 tasks allow for a theory-based assessment of modelling competence, taking into account both atomistic and holistic approaches to defining modelling competence(s)?

RQ 1 is an empirical question to be clarified, while RQ 2 concerns theoretical considerations regarding the potential of comparison tests such as VERA-8.

### 4 Methods

The task dataset under consideration comprises 474 VERA-8 tasks, which are freely available online to registered teachers (<https://www.aufgabenbrowser.de>). According to the website’s filtering function, 168 feature at least one subtask that addresses modelling competence(s). These 168 tasks, which comprise 310 subtasks (266 of which are modelling-focused), were analysed using qualitative content analysis (Kuckartz, 2019). An example is demonstrated in Figure 4.1.

Figure 4.1: Example task “Walking in a Circle”



In order to address RQ 1, an empirical analysis of test items was conducted. The coding framework (see Table 4.1) was based on Maaß’s (2010) classification framework for modelling tasks, which was adapted for the purposes of this study. Key adaptations include the omission of Cognitive demand, the addition of an ‘Other’ category to Situation, the redefinition of the category Openness as Open in problem, Open in solution path, and Open in outcome (Bruder, 2005), and the alignment of Mathematical area with the subject domains of the national standards (KMK, 2004). These alterations were implemented to align with the distinct parameters inherent to the VERA-8 test design. The adapted framework is a tailored, pragmatic approach specifically for VERA-8 tasks. Nevertheless, it can, in principle, be applied to modelling tasks in written tests in general.

Table 4.1: Categories and codes

No.	Categories	Codes
1	Modelling activities	Understand, simplifying & structuring / Mathematising / Working mathematically / Interpreting & validating
2	Data	Matching / Missing / Redundant / Imprecise / Inconsistent
3	Relationship to reality	Authentic & close to reality / Embedded / Intentionally artificial & fantasy
4	Situation	Personal / Occupational / Public / Scientific / Other
5	Type of model	Descriptive / Normative
6	Representations	Text / Table / Picture / Sketch / Diagram / Graph
7	Openness	Open in problem / Open in solution path / Open in outcome
8	Mathematical content	Quantity / Measurement / Space & Shape / Functional relationship / Data & Chance
9	Modelling level	Criticise / Reflect / Select Model

Adopting Maaß’s approach, categories such as Modelling activity (Blum & Leiß, 2007) and Type of model (Meyer, 1984) are deductive categories. Conversely, categories such as Data (Verschaffel et al., 2020) or Relationship to reality function as natural codes. Some categories allow single coding (e.g., Relationship to reality, Type of model, Mathematical content), while others permit multiple coding (e.g., Modelling activities, Data, Situation, Representations, Openness). Coding units were individual subtasks (first cycle coding), later aggregated at task or booklet level (second cycle coding). The coding of the subtasks was

conducted independently by two raters on a binary scale. Disagreements were resolved by consensus in favour of the task (e.g., if raters disagreed on the authenticity of a task, it was rated as authentic). The present procedure was developed to minimise potential negative bias resulting from subjective judgement.

For instance, the task entitled “Walking in a Circle” (see Figure 4.1) was coded as follows: (1) Students must understand the situation and interpret and validate the provided models (i.e., the graphs). (2) Solving the task does not require engagement with numerical data. Nevertheless, one redundant piece of data is redundant (i.e., “one meter”). (3) The context is not necessary, so the task is embedded in reality. (4) The situation described in the task does not correspond to any of the following categories: personal, occupational, public or scientific. Therefore, the situation was coded as ‘Other’. (5) The task involves descriptive modelling and (6) uses graphs for representation. (7) The task was coded as entirely closed. (8) In terms of the mathematical content, students must deal with a functional relationship. (9) They must reflect on the given models and select the correct one.

In order to respond to RQ 2, the empirical findings on RQ 1 were supplemented by theoretical considerations on modelling competence(s) and their assessment, while also taking into account the practical requirements of nationwide comparison tests (Drüke-Noe, 2012).

## 5 Results

### *5.1 Fulfilment of Established Quality Criteria for Modelling Tasks (RQ 1)*

Below, we describe the results chronologically according to the nine categories in Table 4.1.

*1. Modelling Activities:* Only 14 of the 266 modelling-related subtasks require holistic modelling. Among atomistic modelling items, most target mathematising (178 items) and working mathematically (169 items), while understanding, simplifying and structuring (40 items) and interpreting and validating (77 items) appear less often.

*2. Data:* In most cases, tasks provide the exact data needed, so dealing with missing (20 items), imprecise (23 items), or redundant (24 items) information is rarely necessary. Only 2 items simulate data research, such as extracting figures from a table. Items with inconsistent data are absent.

*3. Relationship to reality:* Many tasks (92 items) are essentially mathematical problems with an embedded real-world context. While authentic (80 items) or close-to-reality (85 items) contexts are common, deliberately artificial (5 items) or fictional (2 items) situations occur only sporadically.

*4. Situation:* Most tasks occur in personal (208 items) or occupational (73 items) contexts. Societal (16 items) and scientific (19 items) situations are addressed far less frequently, and some tasks (27 items) cannot be clearly assigned to any of the predefined categories.

*5. Type of model:* Normative modelling only appears in exceptional cases (5 items).

*6. Representations:* Around half of the subtasks (169 items) include at least one

representation, with photographs (52 items) and drawings (55 items) being the most common. Mathematical representations such as diagrams (17 items), coordinate graphs (31 items), or sketches (14 items), as well as tables (39 items) and written materials (6 items), are used less frequently.

7. *Openness of the task*: Openly formulated tasks are the exception: Only a few subtasks allow for openness in the solution process (61 items) or the outcome (35 items); most subtasks (195 items) are fully closed.

8. *Mathematical content*: Of the 266 items, 57 are assigned to the mathematical content quantity, 37 to measurement, 2 to space and shape, 85 to functional relationships, and 86 to data and chance.

9. *Modelling level*: Only a limited number of subtasks (12 items) reach the highest modelling level, requiring students to critically reflect on the choice of model or solution strategy.

Taken together, assuming that modelling competences are best assessed when students engage with tasks meeting defined quality criteria for modelling tasks, the results indicate that VERA-8 may, in many cases, only partially capture students' modelling competence(s). Although all sub-competences are represented, some of them appear relatively infrequently. Missing, imprecise, and redundant data are also uncommon, and a multitude of the tasks lack authenticity. Normative modelling is almost absent, and many tasks are entirely closed.

## ***5.2 Enabling a Theory-based Assessment of Modelling Competence (RQ 2)***

VERA-8 is well suited to the atomistic assessment of modelling competence, although some items also demonstrate that smaller modelling processes can be completed holistically. When participating in VERA-8, teachers receive quantitative results of their participating students at individual or group level (e.g., class or school) on the frequency of correct responses to specific tasks and competences. While qualitative analysis of students' responses is possible, it demands significant effort from teachers, even with the provided support materials. In a computer-based format (already used in some of the 16 German federal states), such analyses could be supported by artificial intelligence. While data from paper-based assessments are static (i.e., coded as correct/incorrect), computer-based formats allow the capture of click patterns, order of task completion, and response times (Hankeln et al., 2025). Comparative assessments are primarily designed as summative instruments, though formative use becomes possible when testing is repeated over time. For example, in the federal state Hamburg (Germany) pupils participate in a comparative assessment on an annual basis. A student ID system enables the tracking of individual competence development (<https://www.kermit-hamburg.de>).

The German educational standards (KMK, 2004) emphasise a component-oriented (atomistic) description of modelling competences. The test items align closely with the steps of the modelling cycle proposed by Blum and Leiß (2007). Concurrently, VERA-8 incorporates subtasks that require holistic modelling, demonstrating that such an approach is, in principle, feasible. To evaluate modelling competence across developmental levels, items reflecting the highest proficiency levels must be included. VERA-8 does, in fact, contain such items, albeit only a few.

## 6 Discussion and Outlook

Regarding RQ 1, the findings suggest limitations in the extent to which VERA-8 tasks meet established quality criteria for modelling tasks. The findings are consistent with similar results reported for tasks used in central examinations in Germany and Austria (Greefrath et al., 2017; Siller & Greefrath, 2020). Regarding RQ 2, the findings show that VERA-8 offers several ways to assess modelling competence(s), which could be further enhanced through computer-based testing, particularly by analysing log files. However, comparative assessments must capture a broad range of competences and meet various design requirements (Drüke-Noe, 2012), meaning certain trade-offs are unavoidable. Nevertheless, the study suggests that the discrepancy between curricular expectations and assessment practice can be reduced in principle, since VERA-8 already contains elements enabling theory-based assessments using both atomistic and holistic approaches to modelling competence(s).

Several criteria should be given greater consideration in future. For example, task texts could contain additional information, or missing data could be simulated by including tables or texts (e.g., newspaper articles). When developing modelling tasks, greater emphasis could also be placed on sub-processes such as Understanding, Simplifying, Structuring, and Interpreting and Validating. Furthermore, more modelling tasks could be developed that align with the mathematical content Measurement as well as Space and Shape.

## References

- 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: Education, Engineering and Economics (ICTMA 12)* (pp. 222–231). Horwood. <https://doi.org/10.1533/9780857099419.5.221>
- Bruder, R. (2005). Working with tasks for the learning of problem solving in maths teaching as an issue of the first teacher training phase. *ZDM*, 37(5), 351–353. <https://doi.org/10.1007/s11858-005-0022-4>
- 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>
- Drüke-Noe, C. (2012). Können Lernstandserhebungen einen Beitrag zur Unterrichtsentwicklung leisten? [Can Learning Assessments Contribute to Instructional Development?] In W. Blum, R. Borromeo Ferri, & K. Maaß (Eds), *Mathematikunterricht im Kontext von Realität, Kultur und Lehrerprofessionalität: Festschrift für Gabriele Kaiser* (pp. 284–293). Vieweg+Teubner. [https://doi.org/10.1007/978-3-8348-2389-2\\_29](https://doi.org/10.1007/978-3-8348-2389-2_29)
- Frejd, P. (2011). An investigation of mathematical modelling in the Swedish national course tests in mathematics. *Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education*, 947–956. <https://hal.science/hal-02158191/>
- 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>
- Greefrath, G. (2015). Problem Solving Methods for Mathematical Modelling. In G. A. Stillman, W. Blum, & M. Salett Biembengut (Eds), *Mathematical Modelling in Education Research and Practice: Cultural, Social and Cognitive Influences* (pp. 173–183). Springer International Publishing. [https://doi.org/10.1007/978-3-319-18272-8\\_13](https://doi.org/10.1007/978-3-319-18272-8_13)
- Greefrath, G., Siller, H.-S., & Ludwig, M. (2017). *Modelling problems in German grammar school leaving examinations (Abitur) – Theory and practice*. 932–939. <https://hal.science/hal-01933483>
- Greer, B., & Verschaffel, L. (2007). Modelling Competencies—Overview. In W. Blum, P. L. Galbraith, H.-

- W. Henn, & M. Niss (Eds), *Modelling and Applications in Mathematics Education: The 14th ICMI Study* (pp. 219–224). Springer US. [https://doi.org/10.1007/978-0-387-29822-1\\_22](https://doi.org/10.1007/978-0-387-29822-1_22)
- Hankeln, C., Adamek, C., & Greefrath, G. (2019). Assessing Sub-competencies of Mathematical Modelling—Development of a New Test Instrument. In G. A. Stillman & J. P. Brown (Eds), *Lines of Inquiry in Mathematical Modelling Research in Education* (pp. 143–160). Springer International Publishing. [https://doi.org/10.1007/978-3-030-14931-4\\_8](https://doi.org/10.1007/978-3-030-14931-4_8)
- Hankeln, C., Kroehne, U., Voss, L., Gross, S., & Prediger, S. (2025). Developing digital formative assessment for deep conceptual learning goals: Which topic-specific research gaps need to be closed? *Educational Technology Research and Development*. <https://doi.org/10.1007/s11423-025-10486-x>
- Henning, H., & Keune, M. (2007). Levels of Modelling Competencies. In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds), *Modelling and Applications in Mathematics Education: The 14th ICMI Study* (pp. 225–232). Springer US. [https://doi.org/10.1007/978-0-387-29822-1\\_23](https://doi.org/10.1007/978-0-387-29822-1_23)
- Kaiser, G. (2007). Modelling and Modelling Competencies in School. In C. Haines, P. Galbraith, W. Blum, & S. Khan (Eds), *Mathematical Modelling ICTMA 12: Education, Engineering and Economics* (pp. 110–119). Horwood. <https://doi.org/10.1533/9780857099419.3.110>
- Kaiser, G. (2020). Mathematical Modelling and Applications in Education. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 553–561). Springer. [https://doi.org/10.1007/978-3-030-15789-0\\_101](https://doi.org/10.1007/978-3-030-15789-0_101)
- KMK. (2004). *Bildungsstandards im Fach Mathematik für den Mittleren Schulabschluss: Beschluss vom 4.12.2003* [Educational Standards in Mathematics for the Intermediate Secondary School Level: Resolution of December 4, 2003]. Luchterhand.
- Krawitz, J., Chang, Y.-P., Yang, K.-L., & Schukajlow, S. (2022). The role of reading comprehension in mathematical modelling: Improving the construction of a real-world model and interest in Germany and Taiwan. *Educational Studies in Mathematics*, 109(2), 337–359. <https://doi.org/10.1007/s10649-021-10058-9>
- Kuckartz, U. (2019). Qualitative Text Analysis: A Systematic Approach. In G. Kaiser & N. Presmeg (Eds), *Compendium for Early Career Researchers in Mathematics Education* (pp. 181–197). Springer International Publishing. [https://doi.org/10.1007/978-3-030-15636-7\\_8](https://doi.org/10.1007/978-3-030-15636-7_8)
- Maaß, K. (2006). What are modelling competencies? *ZDM – Mathematics Education*, 38(2), 113–142. <https://doi.org/10.1007/BF02655885>
- Maaß, K. (2010). Classification Scheme for Modelling Tasks. *Journal Für Mathematik-Didaktik*, 31(2), 285–311. <https://doi.org/10.1007/s13138-010-0010-2>
- Meyer, W. J. (1984). *Concepts of mathematical modeling*. Dover.
- Niss, M., Blum, W., & Galbraith, P. (2007). Introduction. In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds), *Modelling and Applications in Mathematics Education: The 14th ICMI Study* (pp. 3–32). Springer. [https://doi.org/10.1007/978-0-387-29822-1\\_1](https://doi.org/10.1007/978-0-387-29822-1_1)
- Niss, M., & Højgaard, T. (2019). Mathematical competencies revisited. *Educational Studies in Mathematics*, 102(1), 9–28. <https://doi.org/10.1007/s10649-019-09903-9>
- OECD. (2023). *PISA 2022 Assessment and Analytical Framework*. OECD Publishing. <https://doi.org/10.1787/dfc0bf9c-en>
- Schindler, M., Simon, A. L., Baumanns, L., & Lilienthal, A. J. (2025). Eye-tracking research in mathematics and statistics education: Recent developments and future trends. A systematic literature review. *ZDM – Mathematics Education*, 57, 727–743. <https://doi.org/10.1007/s11858-025-01699-8>
- Siller, H.-S., & Greefrath, G. (2020). Modelling Tasks in Central Examinations Based on the Example of Austria. In G. A. Stillman, G. Kaiser, & C. E. Lampen (Eds), *Mathematical Modelling Education and Sense-making* (pp. 383–392). Springer International Publishing. [https://doi.org/10.1007/978-3-030-37673-4\\_33](https://doi.org/10.1007/978-3-030-37673-4_33)
- Verschaffel, L., Schukajlow, S., Star, J., & Van Dooren, W. (2020). Word problems in mathematics education: A survey. *ZDM – Mathematics Education*, 52(1), 1–16. <https://doi.org/10.1007/s11858-020-01130-4>
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds), *Defining and selecting key competencies* (pp. 45–65). Hogrefe & Huber.
- Wirth, L., & Greefrath, G. (2024). Working with an instructional video on mathematical modeling: Upper-secondary students' perceived advantages and challenges. *ZDM – Mathematics Education*, 56(4), 573–587. <https://doi.org/10.1007/s11858-024-01546-2>