

PRE-SERVICE TEACHERS' USE OF CRITERION KNOWLEDGE IN THEIR ANALYSIS OF MODELLING-RELATED CLASSROOM SITUATIONS

Michael Fischer¹, Stefanie Hartmann², Anna Medwed¹, Martin Bracke³,
Sebastian Kuntze²

¹University of Graz, ²Ludwigsburg University of Education, ³University of Koblenz

The use of criterion knowledge related to mathematics education for the interpretation of classroom situations and the quality of learning opportunities is a crucial component of mathematics teacher expertise, hence important for pre-service teacher (PST) education. However, criterion knowledge can be used in different ways and relatively little is known empirically about how and how consistently pre-service teachers use criterion knowledge. This study consequently focuses on this research need by exploring how 253 PSTs used criterion knowledge when analysing modelling-related classroom situations. The study concentrates on the criterion domains of openness and authenticity. The results cast light on contrasts in how criterion knowledge relevant for modelling can be used by PSTs and point to specific professional development needs.

INTRODUCTION

The relevance of mathematical modelling is widely acknowledged (Kaiser, 2020), yet modelling remains a demanding domain of teaching, in particular as it requires teachers' knowledge about task properties and teacher interventions to support students' work without prematurely reducing openness or undermining authenticity through overly sanitised contexts, data, or prescriptions of modelling pathways (Greefrath et al., 2022; Maaß, 2010; Schukajlow et al., 2023).

Research has increasingly focused on how pre-service teachers (PSTs) perceive and judge on classroom situations, which situation aspects they select and attend to in classroom situation episodes, how they interpret them, and which profession-related knowledge they use (Sherin et al., 2011). *Classroom cartoon vignettes*, i.e., representations of practice in cartoon format which invite reflection and analysis (Buchbinder & Kuntze, 2018), are used in teacher education research to elicit such analysis processes (Skilling & Stylianides, 2020). In the present study, we adopt this methodological approach to situations in which mathematical modelling is in the centre, which means that criterion knowledge related to tasks and dealing with tasks, such as openness and authenticity, can be expected to play key roles for situation interpretation. In the context of the project coReflect@maths (coreflect.eu), a classroom cartoon-based vignette instrument was developed and findings related to two mathematising situations indicated that PSTs do not always recognise non-optimal teacher reactions as problematic and sometimes even evaluate them positively (Fischer

et al., 2025; Hartmann et al., 2023). Such situation evaluations might be connected with specific ways how PSTs use their knowledge related to specific criteria for dealing with modelling tasks or tasks in general. However, empirical research on how PSTs use criterion knowledge in their analysis of classroom situations is scarce and therefore a key component of professional expertise remains under-investigated. It is often unclear whether PSTs' judgements are based on a consistent use of modelling task criteria or whether non-criterion-related perceptions prevail in PSTs' analyses.

Consequently, the present study concentrates on PSTs' criteria use in a larger set of ten vignettes: Based on written responses of 253 PSTs to ten cartoon vignettes, we examine how PSTs use task-related criteria to interpret classroom situations. In the following, we outline the theoretical background, derive the research questions, describe design and methods, and report results, which we discuss in a concluding section.

THEORETICAL FRAMEWORK

Modelling-specific pedagogical content knowledge (PCK) can be conceptualised as a bundle of four closely interrelated core categories: knowledge about interventions, modelling processes, modelling tasks, as well as aims and perspectives of modelling (Greefrath et al., 2022). In this paper, we focus on PSTs' criteria use within the task category, as modelling tasks can be described as such through criteria. A criteria-based approach typically foregrounds the aspects of reality, relevance, authenticity, and openness (Maaß, 2010; Schlüter & Besser, 2024; Schukajlow et al., 2023) as key guiding criteria for task development and analysis (Greefrath et al., 2022).

These criteria may appear as binary, but they are multi-faceted: Openness may concern the *initial state* of working on a modelling task (for instance, under-determined tasks are characterised by the fact that “a lot of assumptions have to be made”, Borromeo Ferri, 2018, p. 98), the *intermediate state*, such as methodological openness (Schukajlow et al., 2023), or the *goal state*, i.e., the problem solver also has to specify the question. Likewise, authenticity can relate to the data, the object, the context, the question, and further task elements (Schlüter & Besser, 2024). Conceptually, authenticity in modelling is often distinguished from students' relevance/interest, i.e., a task may be authentic without being experienced as personally relevant (Schlüter & Besser, 2024; Wess et al., 2021). At the same time, the term ‘authenticity’ is used inconsistently in mathematics education and is sometimes employed as a proxy for ‘realistic’ or ‘relevant’, suggesting that PSTs' everyday use may not be congruent to research definitions (Vos, 2011). Moreover, task characteristics can be varied through decisions about, e.g., data provision and contextual simplification. For instance, over-determined tasks may increase solvability but entail “consequences for the authenticity” (Borromeo Ferri, 2018, p. 98). As a consequence, drawing on such criterion knowledge is not merely a technical application of definitions. As shown in prior work, PSTs' evaluations are often shaped by underlying normative orientations (Fischer et al., 2025; Wess et al., 2021). These orientations can act as a filter, leading PSTs to prioritise certain criteria while neglecting others, such as commenting

negatively on open tasks not because they lack quality, but because they conflict with an implicit belief that there should be a single solution pathway or because making assumptions can lead to different possible results (Fischer et al., 2025; Schukajlow et al., 2023). Accordingly, examining how PSTs use modelling task criteria can help to clarify how PCK is applied when making judgements about tasks or teacher support.

Greefrath et al. (2022) stress that theoretical PCK does not automatically ensure an understanding of learners' difficulties and that "to this end, practical or working experience is required" (p. 388). This provides a rationale for using vignette-based approximations of practice in PST education and related research: Vignette analyses can offer repeated, comparable opportunities to use modelling-related criteria, such as openness and authenticity, when interpreting tasks and teacher (re)actions. This process is a component of *teacher noticing*, which captures how teachers selectively attend to classroom events or situation aspects and interpret them through professional knowledge-based reasoning (Sherin et al., 2011). In modelling contexts, noticing can be considered particularly demanding as teachers do not only have to monitor mathematical correctness but they also have to interpret students' modelling processes and respond in ways that maintain student independence and decision-making (Kaiser, 2020; Schukajlow et al., 2023). Findings from modelling-specific noticing research show that PSTs' perception of students' difficulties varies substantially depending on the type of difficulty encountered in modelling situations (Alwast & Vorhölter, 2022) and Paul et al. (2023) highlight that experts' individual noticing focuses regarding task potential are not always consistent across different vignettes.

Research is therefore needed into how PSTs' noticing can draw on criteria: judgements about instructional quality may depend on whether and how PSTs use modelling-related criteria when evaluating both the task and the vignette teacher's support for students' work on those tasks. For such research, vignettes can provide an approach to eliciting and comparing such PST analyses under controlled conditions (Buchbinder & Kuntze, 2018; Skilling & Stylianides, 2020).

RESEARCH QUESTIONS

If theoretical PCK does not automatically translate into an understanding of learners' difficulties, practice-proximal formats such as classroom vignettes are useful to investigate how PSTs use professional knowledge in situation interpretation. Against this background, we address the following research questions: *How do PSTs use the criteria of openness and authenticity in their written analyses of modelling-related classroom cartoon vignettes? To what extent is this criteria use vignette-specific, and to what extent does it recur across multiple vignettes for the same PST?*

METHODS

The present study is a secondary analysis extending prior vignette based work on PSTs' analyses of non-optimal modelling support (Fischer et al., 2025; Hartmann et al., 2023). The data set consists of written paper-and-pencil responses in which 253 PSTs

from three German-speaking universities (in Austria and Germany) provided open evaluations of situations represented in a set of ten classroom cartoon vignettes. In line with the research questions, we aggregate data across educational backgrounds and study progress to examine patterns of criteria use across the vignette set. The vignettes make non-optimal forms of modelling support deliberately visible to prompt criteria-based analysis, enabling us to examine whether and how PSTs use task-related criteria, and to what extent criterion use is vignette-specific versus recurring across multiple situations for individual PSTs.

To give a brief sense of the classroom situations captured by the ten vignettes (V1–V10), and to clarify later references in the results, we outline three examples. In V8, the vignette students are asked to estimate the number of people in a traffic jam that actually occurred (see also Wess et al., 2021), and the vignette teacher fails to invite reflection on students' assumptions and modelling decisions. Another vignette (V1) involves estimating the number of noodles in a package similar to a Fermi-type task (Borromeo Ferri, 2018), but the vignette teacher prematurely prescribes a weight-based approach using a scale, thereby narrowing methodological openness (Schukajlow et al., 2023). Similarly, in vignette V2 the teacher asks how many steps it would take to get to Paris, with an atlas provided as a contextual resource, yet already restricts to a single method while posing the task.

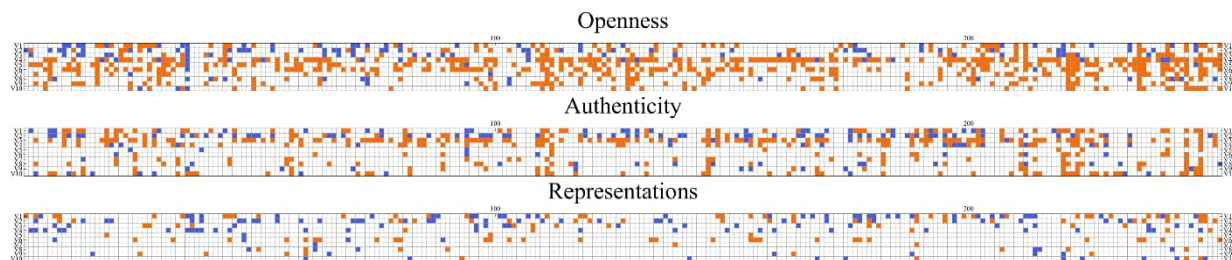
Coding followed theory-based categories (Borromeo Ferri, 2018; Schlüter & Besser, 2024; Schukajlow et al., 2023; Wess et al., 2021). Each vignette response was coded as 1 when a PST used the respective criterion in a theory-aligned way, that is, when openness or authenticity was mobilised as a quality criterion in line with modelling-task research (this may involve praising openness/authenticity when present or criticising its absence). Responses were coded as -1 when the criterion was invoked in a way that contradicts these theoretical accounts, and as 0 when no discernible reference occurred. As *authenticity* is conceptually multifaceted and may overlap with realism/relevance in everyday discourse (Vos, 2011), we treated those references as part of authenticity-related reasoning. Accordingly, we coded authenticity whenever PSTs framed the vignette task/situation as (non-)genuine and/or as (non-)relevant for learners, without requiring terminological precision. In addition, we coded whether PSTs referred to the representations provided in the vignettes (e.g., a noodle package, an atlas) as part of their authenticity reasoning, distinguishing such object-related references from broader claims about the authenticity of contexts, data, or questions (Schlüter & Besser, 2024).

The coding was carried out by two authors independently and yielded good inter-rater reliability (overall $\kappa = 0.749$). To visualise criteria-use profiles, all codes were organised in a person-by-vignette matrix, which allows two complementary readings: horizontal cuts describe vignette-specific patterns across PSTs, and vertical cuts capture repeated criteria use across vignettes for the same PST.

RESULTS

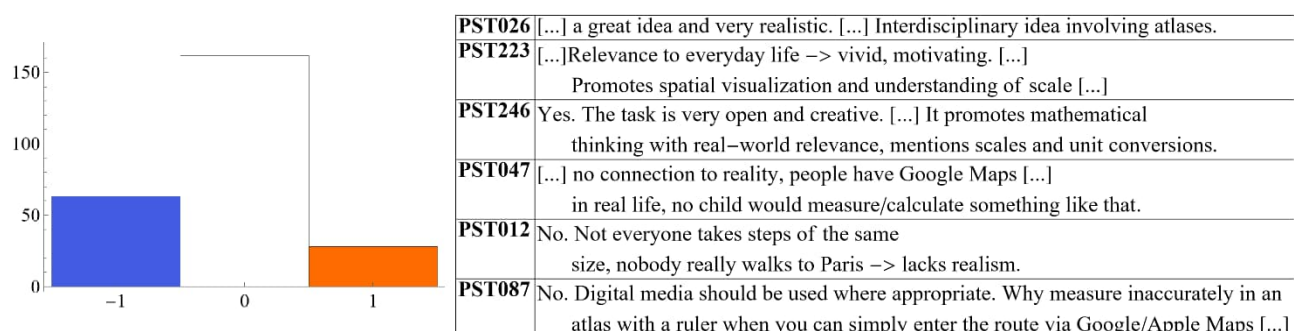
Figure 1 visualises criteria use across the full data set. In the matrix, orange indicates a theory-aligned use of a criterion, white indicates no discernible reference, and blue indicates a theory-misaligned use of the criterion (codes 1, 0, -1). The figure shows clear differences in criteria use both vertically, across the ten vignettes within individual PSTs, and horizontally, across PSTs within individual vignettes.

Figure 1: Criteria use of 253 PSTs across 10 vignettes.



From a horizontal perspective, Figure 1 shows that V5 and V6 elicited frequent openness-related reasoning, although openness was also taken up in V1–V4 and was used systematically by some PSTs across most of the vignette set. V1–V4, all centred on how a modelling task is posed, also elicited frequent authenticity-related reasoning, including both theory-aligned and theory-misaligned references. V2 illustrates this split (Figure 2): several PSTs framed the task in V2 as authentic, realistic, and relevant, for example describing it as “a great idea and very realistic”, emphasising its “relevance to everyday life”, or highlighting that it is “very open and creative”. At the same time, other PSTs questioned the legitimacy of the task, for example stating that there is “no connection to reality”, arguing that “nobody really walks to Paris”, or criticising the prompt as non-authentic because one would not “measure inaccurately in an atlas” when digital maps are available.

Figure 2: Selected authenticity criteria use in V2.



These opposing authenticity framings were often connected to evaluations of representations provided in the task (Figure 1, bottom). Positive comments treated, for example, the atlas as a productive real-world anchor or an interdisciplinary resource. Critical comments, in contrast, interpreted the atlas as an artificial add-on or an

outdated tool, thereby questioning authenticity and, in some cases, linking this critique to concerns about methodological openness. Similar representation-sensitive authenticity-related reasoning reappears in V3 and V4. V3, about a dripping water tap, again led many PSTs to judge whether the context and data warrant modelling or are perceived as contrived; several responses also stressed demands for exactness when working with assumptions. Across several vignettes with under-determined initial states, i.e., situations that require estimation, assumption-making, and working with variable data, some PSTs evaluated these features critically and called for more explicit guidance (e.g., given measures, tools, or fixed starting points).

We now turn to vertical profiles, shifting from vignette-level comparisons to within-person patterns across the ten vignettes. We illustrate the profiles found in the analysis with two contrasting cases (Figure 3), which exemplify the person-level variation suggested by Figure 1. PST035 (Fig. 3, left) showed repeated openness-related reasoning across the vignette set, but consistently invoked openness in theory-misaligned ways, framing under-determination as a deficit. For instance, PST035 criticised tasks as vague or underspecified (e.g., “too imprecise”, “not clear what the teacher wants”) and repeatedly demanded that measures, tools, or fixed starting points be provided by the teacher.

PST212 (Fig. 3, right) also referred to openness repeatedly, yet did so in a theoretically-aligned way, emphasising that students should “model on their own” rather than follow a prescribed method. This openness use did not translate into affirmative authenticity judgements; instead, PST212 frequently questioned tasks’ real-world relevance “for students”, often denying subjective relevance and associating representations with hindering realistic enactment (e.g., feasibility issues or the work with a pack of noodles being food waste). Together, these two cases form contrasting criteria-use profiles (Figure 3), illustrating how openness- and authenticity-related reasoning can appear repeatedly across multiple vignettes within the case of the same PST.

Figure 3: Selected abbreviated results on criteria use for authenticity, openness, and the use of representations from two PSTs.

PST035					
V1	O	A	R	V1	[...] would be too imprecise for me, since each bag could contain a different number of pasta, because they are weighed rather than counted [...]
V2				V2	[...] depending on the age, I would have given the students some of the measurements—or all of them.
V3				V3	[...] I like the structure of the [pre-structured] worksheet [...]
V4				V4	[...] because here it is not clear what the teacher wants. [...] I would have guided the situation more.
V5				V5	[...] unsuitable for younger children, since they cannot yet estimate speed [...]
V6				V6	[...] paint buckets state approximately how many m ² they cover; otherwise, it is a nice, realistic task.
V7				V7	
V8				V8	
V9				V9	
V10				V10	

PST212					
V1	O	A	R	V1	[...] students see no reason to engage [...] 500g of pasta as food waste [...] Disadvantages: students do not do the modelling themselves
V2				V2	[...] only too little real-world relevance for students [...] model is not developed by students
V3				V3	
V4				V4	
V5				V5	
V6				V6	negative – no real-world relevance. [...] does not let the students model themselves → imposes his solution path.
V7				V7	
V8				V8	[...] Negative – no real-world relevance for students [...] Positive: [...] lets them [independently] model/calculate → good [...]
V9				V9	
V10				V10	Positive: a realistic model, a comprehension-check task [...]

DISCUSSION AND CONCLUSION

This study examined PSTs’ criteria use regarding openness and authenticity when analysing classroom cartoon vignettes, and whether such criteria use is mainly vignette-specific or appears repeatedly across the vignette set. Considering the full data

set, the results show pronounced heterogeneity both horizontally and vertically: the same vignette features prompted divergent judgements across PSTs, particularly in authenticity-related criterion use. Moreover, the findings suggest that PSTs draw on different aspects of authenticity, shifting between authenticity of representations, data, and contexts and the subjective relevance of the question itself (Schlüter & Besser, 2024; Vos, 2011), as visible in profiles such as PST212. Vertically, the profiles indicate that criteria-related reasoning may either appear repeatedly across multiple vignettes within the same PST or remain more episodic and vignette-bound. Taken together, these patterns support the view that criteria use is a meaningful indicator of how PSTs interpret modelling tasks and teacher support across a range of modelling situations.

Importantly, the results also help to relate criteria use to broader normative orientations in modelling. In our data, theory-misaligned uses of openness and authenticity were particularly common in under-determined modelling situations requiring estimation, assumption-making, and working with variable, authentic data. One interpretation is that some PSTs approach such situations through a precision-oriented view of mathematics, where demands for exactness can outweigh modelling-specific practices of validation and plausibility under uncertainty (Borromeo Ferri, 2018; Wess et al., 2021). In line with this, some PSTs treated under-determination as a deficit that calls for stronger teacher direction, rather than as a productive feature of modelling tasks.

Two limitations should be acknowledged. First, the data stem from cartoon vignettes and written responses, which support comparability but do not capture in-the-moment classroom interaction or actual teaching behaviour (Buchbinder & Kuntze, 2018). Second, the sample is not intended to be representative, and findings may depend on the specific vignette set and its design choices; they should therefore be interpreted as evidence of identifiable patterns of criteria use within this instrument rather than as generalisable prevalence estimates for larger PST populations.

Outlook for follow-up work is straightforward. Empirically, the next step is to formalise the profiling approach by clustering person-by-vignette matrices and examining how cluster membership relates to evaluations of non-optimal teacher support (Hartmann et al., 2023; Paul et al., 2023). For PST education, the findings underline the need for focused learning opportunities that explicitly address openness and authenticity in modelling tasks. Classroom cartoon vignettes can provide practice-proximal cases for seminars on posing modelling tasks, discussing their properties and ways of working with these tasks in the classroom. Such vignettes offer repeated opportunities to articulate, challenge, and refine criteria-based judgements in a structured setting that remains relevant for classroom practice.

References

- Alwast, A., & Vorhölter, K. (2022). Noticing students' difficulties while working on modelling problems. In C. Fernández, S. Llinares, A. Gutiérrez, & N. Planas (Eds), *Proceedings of the 45th Conf. of the IGPME* (Vol. 4, p. 168). PME.

- Borromeo Ferri, R. (2018). *Learning How to Teach Mathematical Modeling in School and Teacher Education*. Springer International Publishing.
- Buchbinder, O., & Kuntze, S. (Eds). (2018). *Mathematics Teachers Engaging with Representations of Practice*. Springer International Publishing.
- Fischer, M., Bracke, M., Hartmann, S., & Kuntze, S. (2025). Teachers' tacit assumptions in their analysis of non-optimal modelling support: A vignette-based study. In C. Cornejo, P. Felmer, D. M. Gómez, P. Dartnell, P. Araya, A. Peri, & V. Randolph (Eds), *Proceedings of the 48th Conf. of the IGPME* (Vol. 1, pp. 251–258). PME.
- Greefrath, G., Siller, H.-S., Klock, H., & Wess, R. (2022). Pre-service secondary teachers' pedagogical content knowledge for the teaching of mathematical modelling. *Educational Studies in Mathematics*, 109(2), 383–407. <https://doi.org/10.1007/s10649-021-10038-z>
- Hartmann, S., Krummenauer, J., & Kuntze, S. (2023). Mathematics teachers' analysis of modelling situations in the classroom: A vignette-based study. *Thirteenth Congress of the European Society for Research in Mathematics Education (CERME13)*, 14.
- Kaiser, G. (2020). Mathematical Modelling and Applications in Education. In S. Lerman (Ed.), *Encyclopedia of Mathematics Education* (pp. 553–561). Springer International Publishing. https://doi.org/10.1007/978-3-030-15789-0_101
- 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>
- Paul, J. F., Dreher, A., Wang, T.-Y., Hsieh, F.-J., Hansen, L., & Lindmeier, A. (2023). Are experts' noticing focuses regarding the learning potential of tasks and its use consistent across instructional situations? A secondary analysis. In M. Ayalon, B. Koichu, R. Leikin, L. Rubel, & M. Tabach (Eds), *Proceedings of the 46th Conf. of the IGPME* (Vol. 4, pp. 27–34). PME.
- Schlüter, D., & Besser, M. (2024). Assessing authenticity in modelling test items: Deriving a theoretical model. *Frontiers in Education*, 9.
- Schukajlow, S., Krawitz, J., Kanefke, J., Blum, W., & Rakoczy, K. (2023). Open modelling problems: Cognitive barriers and instructional prompts. *Educational Studies in Mathematics*, 114(3), 417–438. <https://doi.org/10.1007/s10649-023-10265-6>
- Sherin, M., Jacobs, V. R., & Philipp, R. A. (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. Taylor & Francis.
- Skilling, K., & Stylianides, G. J. (2020). Using vignettes in educational research: A framework for vignette construction. *International Journal of Research & Method in Education*, 43(5), 541–556. <https://doi.org/10.1080/1743727X.2019.1704243>
- Vos, P. (2011). What Is 'Authentic' in the Teaching and Learning of Mathematical Modelling? In G. Kaiser, W. Blum, R. Borromeo Ferri, & G. Stillman (Eds), *Trends in Teaching and Learning of Mathematical Modelling* (pp. 713–722). Springer Netherlands.
- Wess, R., Klock, H., Siller, H.-S., & Greefrath, G. (2021). *Measuring Professional Competence for the Teaching of Mathematical Modelling: A Test Instrument*. Springer International Publishing.