DEVELOPING A QUESTIONNAIRE ON QUALITY PRINCIPLES IN PRIMARY MATHEMATICS TEACHING: INCORPORATING STUDENT AND EXPERT PERSPECTIVES

Sophia Rothering[1], Bettina Roesken-Winter[1], Malte Lehmann[1], Lars Meyer-Jenßen[2]

[1]University of Münster, [2]Health and Medical University Erfurt

Although student ratings are increasingly used to evaluate teaching quality, prior research has focused primarily on teaching in general or on secondary school students. This study therefore developed a questionnaire to capture primary school students' perspectives on mathematics teaching. The process involved two phases of cognitive interviews with 25 primary school students and one phase of expert ratings provided by ten experts in mathematics education. The resulting 20-item questionnaire was broadly comprehensible to the students and covered the intended constructs, highlighting the importance of integrating both student and expert perspectives in a unified procedure for developing comprehensible, theoretically grounded items.

INTRODUCTION

Teaching quality is a key determinant of student learning outcomes (Hattie, 2008) and can be evaluated using various methods. Although classroom observations conducted by trained external experts are a common method, they are time-consuming and difficult to implement on a large scale, especially when multiple observations are required for reliability (Hill et al., 2012). In recent years, student ratings have become increasingly relevant for assessing teaching quality worldwide (Bijlsma, 2021). Studies have shown that students can provide reliable and valid feedback and that their perceptions are linked to relevant educational outcomes, such as academic achievement (Fauth et al., 2014; Wagner et al., 2013). Therefore, student feedback has the potential to not only provide information for improving teaching, but also to empower students to participate in shaping their learning environment. However, most existing studies focus on either general teaching quality or feedback from secondary school students (e.g., Liu et al., 2024; van der Lans et al., 2021). Research into specific subjects, particularly in primary education, remains limited, despite evidence that even young children can offer meaningful evaluations when items are developmentally appropriate (van der Scheer et al., 2019).

The goal of our study is to address this gap by developing a questionnaire that captures primary school students' perceptions of quality principles in mathematics education. Due to the developmental stage of primary school students, particular emphasis is placed on content comprehensibility to ensure that items are interpreted as intended. Furthermore, content validity is supported by expert ratings to verify that the items adequately capture the intended constructs (Messick, 1995).

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. 1235-1242. PME.

THEORETICAL BACKGROUND

Student Feedback on Teaching Quality

The importance of students assessing teaching quality from their perspective has increased considerably in recent years. Research has shown that questionnaires are an effective tool to evaluate teaching quality in both primary and secondary education (Bijlsma, 2021). Moreover, student feedback is a valuable resource for teachers' professional development, as students' firsthand experience of the learning environment enables them to provide insights into classroom processes that external observers might overlook (Fauth et al., 2014). Studies have also consistently demonstrated that student ratings provide meaningful and differentiated information. For example, Wagner et al. (2013) and Fauth et al. (2014) found that even young students can differentiate between various aspects of teaching quality. Further research has shown that student feedback reflects important teaching aspects and can indirectly indicate learning outcomes (Liu et al., 2024). Although biases such as the halo effect may occur due to students' sense of communion, they can still provide valuable feedback when carefully interpreted (Röhl & Rollett, 2021).

Quality Principles for Mathematics Teaching

Examining existing instruments reveals significant variation in how teaching quality is defined. One widely cited framework is the three-dimensional model by Klieme et al. (2009), which includes supportive climate, cognitive activation and classroom management. Many authors have expanded upon this model by adding dimensions such as student-oriented instruction. However, some argue that the framework is too general and does not sufficiently capture the specific characteristics of teaching in different subject areas (Liu et al., 2024). To examine teaching quality in mathematics education more specifically, several studies have highlighted mathematics-specific quality principles. Prediger et al. (2022) synthesized these into five principles for high-quality mathematics teaching: *Conceptual Focus*, *Cognitive Demand*, *Enhanced Communication*, *Student Focus and Adaptivity*, and *Longitudinal Coherence.*

The principle of *Conceptual Focus* emphasizes that mathematics teaching should foster students' conceptual understanding and meaningful connections between mathematical procedures and underlying concepts (Hiebert & Carpenter, 1992). Furthermore, the principle of *Cognitive Demand* is incorporated, emphasizing that tasks and problems should stimulate higher-order thinking processes (Renkl, 2015). The third principle of *Enhanced Communication* supports and maintains the initial two principles by encouraging student communication for learning and learning through communication. This approach fosters rich mathematical discussions in both small groups and whole-class settings, where students can develop mathematical concepts and practices (Walshaw & Anthony, 2008). As Liu et al. (2024) argued, additional principles are necessary to comprehensively describe high-quality mathematics teaching. Accordingly, the principle of *Student Focus and Adaptivity* emphasizes

monitoring students' learning progress and adapting instruction to promote their thinking towards formal mathematical understanding. Furthermore, the principle of *Longitudinal Coherence* pertains to organizing subject matter content along long-term learning trajectories within a coherent spiral curriculum (Bruner, 1966). Together, these five principles form the conceptual basis of the ten-year, national professional development program, QuaMath. They are referred to as the QuaMath principles in the following (Prediger et al., 2022).

PRESENT STUDY AND RESEARCH QUESTION

This study aims to develop a questionnaire that measures primary students' perceptions of the quality of mathematics teaching, guided by the QuaMath principles. According to AERA, APA and NCME (2014), the development process should include gathering evidence from different categories for testing validity. The primary focus is on the response processes of students to understand how they interpret the items and how their responses offer insight into the clarity of the questionnaire. Additionally, analyzing the experts' responses will provide further insight into the questionnaire's consistency with the five QuaMath principles in terms of content validity. Accordingly, the current study addresses the following research questions:

RQ1: How do primary school students comprehend and interpret the items designed to assess their perceptions of teaching quality in terms of the QuaMath principles of high-quality mathematics teaching?

RQ2: How do experts evaluate the clarity and alignment of the items with the QuaMath principles of high-quality mathematics teaching?

METHODOLOGY

Participants, Context, and Data Collection

A total of 25 second- and third-grade students from ten primary schools participated in the cognitive interview study. The children had an average age of 8.2 years (SD = 0.8), with 14 boys and 11 girls taking part. Ten experts in mathematics education, averaging 11.2 years of professional experience (SD = 3.8), participated in an expert rating. The experts reported high familiarity with the QuaMath principles (M = 9.1, SD = 0.7) and moderate familiarity with questionnaire construction (M = 6.4, SD = 2.7), with both scores on a scale from one to ten.

The questionnaire was developed to operationalize the QuaMath principles from the students' perspective. Initial items were derived directly from theoretical descriptions of the principles using a deductive approach. Existing questionnaires addressing related constructs, such as the instrument developed by Fauth et al. (2014), were also consulted to ensure conceptual alignment and to draw on established item formats. According to Bell (2007), the items were simplified for primary school students by reducing their syntactic complexity and avoiding abstract terminology as well as negative formulations. A four-point scale was implemented to capture students' responses

regarding their mathematics lessons during the previous week. Response options ranged from “in every lesson” to “in no lesson”.

Data collection for RQ1 involved conducting one-on-one cognitive interviews. During each interview, students read each item aloud and selected their responses. Then, they were asked to explain their choices. If more information was needed to evaluate their comprehension, prompts were used to encourage the students to rephrase the sentence in their own words, provide an example, or explain specific terms. All interviews were audio recorded and transcribed for analysis. Data collection for RQ2 consisted of expert ratings. Each expert was presented with a set of items related to the five QuaMath principles, and was asked to select the five most suitable items per principle and provide justifications for their choices. To identify potentially missing content in the questionnaire, experts were further asked to comment on each principle to indicate whether any aspects of the principles were missing or insufficiently represented.

The development and testing of the items followed three phases. In the first phase, nine cognitive interviews were conducted using an initial set of ten items representing two of the five principles. The goal was to determine whether students could comprehend longer sentence structures and mathematics-specific terminology. The results informed the revision of the initial items and the development of additional items for the remaining three principles. Subsequently, six interviews were conducted using the expanded item pool to ensure the comprehensibility of the newly constructed items for students. In the second phase, experts selected the five most suitable items for each principle and noted any missing content. Based on these estimations, a final set of items was created. A third phase included eleven more cognitive interviews with this set of items to confirm that they were interpreted as intended. These interviews enabled final adjustments to be made to the questionnaire.

Data Analysis

To answer RQ1, we analyzed the transcribed cognitive interviews using a structured, content-focused approach. For each item, we determined the extent to which students’ explanations aligned with the intended meaning, as defined by the five QuaMath principles. Each response was coded as either “completely correct” (1) or “incorrect or partially correct” (0), providing a systematic measure of interpretive accuracy. To ensure coding reliability, two researchers independently coded all interview data, and clarified any discrepancies in a structured consensus process. Coding was conducted across all development phases to identify recurring patterns of misunderstanding, ambiguous wording, and insufficient conceptual alignment. Descriptive statistics were then used to summarize the frequencies of correct and incorrect interpretations per item to provide insights into how students interpreted the items.

To answer RQ2, we coded each item from the expert ratings as either “selected” (1) or “not selected” (0). Since each expert was required to select exactly five items per principle, the resulting binary data reflected the perceived relative relevance of each

item rather than an absolute judgement of quality. For each item, we summed the number of selections made by all experts, yielding an endorsement frequency ranging from zero to nine. Items with higher frequencies were interpreted as having stronger content validity and clearer alignment with the intended constructs. Based on these frequencies, we formed three categories. Items receiving between six and nine votes were classified as high-consensus items. Items with three to five votes were categorized as moderate-consensus items and considered candidates for revision or removal, whereas items with zero to two votes were classified as low-consensus items and subsequently removed. Additionally, we examined the qualitative comments provided by the experts to identify missing aspects of the constructs and to guide both the inclusion of new items and the content-related and linguistic modification of existing items.

RESULTS

As mentioned in the methodology section, the questionnaire was developed in three phases. In the first phase, cognitive interviews were conducted with students to create an initial pool of items. Phase two involved experts rating these items. The results of the first two phases led to a final round of cognitive interviews with students in phase three. To answer RQ1, we first examined how students interpreted the items in phase one. Several comprehension issues related to wording and sentence structure emerged during the cognitive interviews. Some of these difficulties stemmed from abstract or ambiguous terms such as “idea”, “pattern” and “similar”. Misunderstandings also occurred at the sentence level due to complex sentence structures and terms being interpreted in unintended contexts. For instance, the phrase “words and sentences” caused students to associate the item with German lessons instead of mathematics, which led to confusion about its intended meaning:

Luis:	*(reads the item)* I learn words and sentences, that I need in math. So, whether I sometimes use sentences for that?
Interviewer:	Whether you learn words and sentences in math.
Luis:	That’s not German lessons. I think in German lessons you learn more sentences.

This example illustrates the importance of linguistic clarity in enabling students to reliably understand and respond to items. Accordingly, the items were revised by simplifying sentence structures, specifying abstract terms, and removing expressions tending to contextual misinterpretation.

After revising the questionnaire and omitting the principle of *Longitudinal Coherence* as suggested by the expert ratings (see Results for RQ2), the final 20-item version was tested in a third phase of cognitive interviews with second- and third-grade students. Table 1 provides an overview of the proportion of correct item interpretations.

Table 1: Proportion of correct interpretations of the items of four of the QuaMath of high-quality mathematics teaching in the final round of cognitive interviews

	2nd grade	3rd grade	Average
Conceptual Focus	60%	93%	78%
Enhanced Communication	50%	80%	66%
Cognitive Demand	60%	96%	80%
Student Focus and Adaptivity	93%	96%	95%

The comprehension rate was satisfactory on average, ranging from 78% to 95% for three of the principles. However, students' correct interpretation of the *Enhanced Communication* items was slightly lower. Third-grade students demonstrated higher comprehension overall than second-grade students. While the instrument was generally understandable, the results suggest that it is particularly suitable for students in third grade and older.

Regarding RQ2, the expert ratings provided further insight into the quality of the item pool. As shown in Table 2, the number of expert endorsements varied by principle. Some constructs displayed strong consensus, while others demonstrated more heterogeneous evaluations.

Table 2: Distribution of chosen items of the QuaMath principles of high-quality mathematics teaching in the expert ratings

	Items with …		
	high consensus	moderate consensus	low consensus
Conceptual Focus	2	8	7
Enhanced Communication	4	6	4
Cognitive Demand	4	5	4
Student Focus and Adaptivity	2	7	4
Longitudinal Coherence	6	1	3

Analysis of the vote distribution revealed significant variation in the frequency with which individual items were selected. Notably, the distribution of expert endorsements for the principles of *Conceptual Focus* and *Student Focus and Adaptivity* indicated a clear need for further refinement. Although several items received moderate ratings, the overall patterns suggested that the existing formulations did not fully capture the intended conceptual emphases. In a subsequent step, the additional qualitative

comments provided by the experts were systematically analyzed to further refine the instrument and adding further items. These comments provided important insights into aspects, that had not been fully addressed by the rating distributions alone. According to the principle of *Longitudinal Coherence*, the expert ratings showed that the items accurately reflected the principle in terms of content. However, the experts described the principle as conceptually challenging and not meaningfully measurable by students. Consequently, it was excluded from the instrument. Further revisions were prompted by comments noting that several items placed an excessively strong emphasis on arithmetic, which risked narrowing the intended conceptual scope. Additionally, certain terms, such as "understanding" were perceived as overly abstract and were therefore reworded to enhance clarity and accessibility for students.

CONCLUSION

Overall, the findings provided valuable insights into the challenges of creating items for primary students and emphasized the importance of iterative testing. By combining the perspectives of students and experts, we were able to revise the items, enhancing the clarity, conceptual precision, and alignment with the core aspects of each principle. Thus, integrating these two perspectives in three phases offered valuable insights for creating understandable, construct-based items. Given the students' young age and corresponding abilities, it was helpful to begin and end with the student perspective, while also incorporating the experts' views. To validate the questionnaire further, it will be administered to entire classrooms to assess its applicability in authentic classroom situations and analyze its construct validity and reliability. Ultimately, the questionnaire may provide formative feedback to help teachers improve their mathematics lessons and give learners a meaningful voice in the evaluation process.

Additional Information: This research was funded within the project QuaMath, financed by the Standing Conference of the Ministers of Education and Cultural Affairs (KMK) in Germany.

References

American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME]. (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Bell, A. (2007). Designing and testing questionnaires for children. *Journal of Research in Nursing 12*(5), 461–469.

Bijlsma, H. (2021). The quality of student perceptions questionnaires: A systematic review. In W. Rollett, H. Bijlsma, & S. Röhl (Eds.), *Student feedback on teaching in schools: Using student perceptions for the development of teaching and teachers* (pp. 47–72). Springer.

Bruner, J. (1966). *Toward a theory of instruction*. Harvard University Press.

Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school. Dimensions and prediction of student outcomes. *Learning and Instruction, 29*, 1–9.

Hattie, J. (2008). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement.* Routledge.

Hiebert, J., & Carpenter, T. P. (1992). Learning and teaching with understanding. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 65–97). Macmillan.

Hill, H. C., Charalambous, C. Y., & Kraft, M. A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher, 41*(2), 56–64.

Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study: Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Seides (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137–160). Waxmann.

Liu, X., Yang Hansen, K., De Neve, J., & Valcke, M. (2024). Teacher versus student perspectives on instructional quality in mathematics education across countries. *Instructional Science, 52,* 477–513.

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.

Prediger, S., Götze, D., Holzäpfel, L., Rösken-Winter, B., & Selter, C. (2022). Five principles for high quality mathematics teaching: Combining normative, epistemological, empirical, and pragmatic perspectives for specifying the content of professional development. *Frontiers in Education, 7*(969212), 1–15.

Renkl, A. (2015). Different roads led to Rome: the case of principle-based cognitive skills. *Learning: Research and Practice, 1*(1), 79–90.

Röhl, S., & Rollett, W. (2021). Student perceptions of teaching quality: Dimensionality and halo effects. In W. Rollett, H. Bijlsma, & S. Röhl (Eds.), *Student feedback on teaching in schools: Using student perceptions for the development of teaching and teachers* (pp. 31–46). Springer.

Wagner, W., Göllner, R., Helmke, A., Trautwein, U., & Lüdtke, O. (2013). Construct validity of student perceptions of instructional quality is high, but not perfect: Dimensionality and generizability of domain-independent assessments. *Learning and Instruction, 28*, 1–11.

Walshaw, M., & Anthony, G. (2008). The teacher's role in classroom discourse: A review of recent research into mathematics classrooms. *Review of Educational Research, 78*(3), 516–551.

Van der Lans, R. M., Maulana, R., Helms-Lorenz, M., Fernándenz-García, C.-M., Chun, S., de Jager, T., Irnidayanti, Y., Inda-Caro, M., Lee, O., Coetzee, T., Fadhilah, N., Jeon, M., & Moorer, P. (2021). Student Perceptions of Teaching Quality in Five Countries: A Partial Credit Model Approach to Assess Measurement Invariance. *Sage Open, 11*(3), 1–20.

Van der Scheer, E. A., Bijlsma, H. J. E., & Glas, C. A. W. (2019). Validity and reliability of student perceptions of teaching quality in primary education. *School Effectiveness and School Improvement, 30*(1), 30–50.