Developing and Piloting a Responsive Mathematics Teaching Observation Scale (RMTOS)

Zara Ersozlu
The University of Newcastle
zara.ersozlu@newcastle.edu.au

This study reports the development and piloting of the Responsive Mathematics Teaching Observation Scale (RMTOS), a structured classroom observation instrument designed to capture teachers' responsiveness to students' mathematical thinking. We trialled the instrument across 24 mathematics lessons observed twice weekly over six weeks across two Year 6 classrooms in a Catholic primary school. The findings provide preliminary evidence that responsive mathematics teaching can be operationalised through structured observation, and the refined instrument offers a promising tool for future research on mathematics instruction and teacher professional development.

Introduction

Responsive teaching in mathematics has become an important focus in mathematics education because it foregrounds teachers' attention to students' ideas and their capacity to adapt instruction in response to those ideas. Rather than treating teaching as the delivery of predetermined procedures, responsive teaching emphasises noticing the substance of students' mathematical thinking, interpreting the significance of that thinking, and using it to shape subsequent instructional moves (Jacobs et al., 2010; Sherin et al., 2011; Robertson et al., 2016). In mathematics classrooms, such responsiveness is especially important because students' explanations, strategies, partial understandings, and misconceptions provide critical resources for advancing conceptual understanding and collective sense-making (Kazemi & Stipek, 2001; Lampert, 2001). The aim of the study, along with the guiding research questions, is presented following the literature review.

Literature Review

Responsive teaching is defined here as teachers' capacity to notice, interpret, and build on students' mathematical thinking during instruction, drawing on research on professional noticing in which teachers recognise students' strategies, interpret their mathematical meanings, and decide how to respond instructionally (Jacobs et al., 2010; Robertson et al., 2016; Sherin et al., 2011). This conception aligns with scholarship on ambitious and discussion-based mathematics teaching, where classroom discourse is organised not simply to elicit answers, but to make reasoning public, compare strategies, and press for explanation and justification (Chapin et al., 2009; Lampert, 2001; Kazemi & Stipek, 2001). In this sense, managing cognitive load is also important, as it helps teachers reduce unnecessary task demands so that students can focus more fully on understanding, reasoning, and explaining their mathematical thinking (Sweller, 2024; van Gog et al., 2011). Across this body of work, responsive teaching is understood as an interactional practice in which teaching decisions are shaped by students' developing ideas rather than by a fixed script (Robertson et al., 2016). It is important to note that responsive mathematics teaching, as theorised in this tradition, is conceptually distinct from culturally responsive pedagogy (Gay, 2010) and culturally relevant pedagogy (Ladson-Billings, 1995). Whereas culturally responsive pedagogy concerns the use of students' cultural knowledge and experiences as resources for learning, responsive mathematics teaching concerns the use of students' content-specific mathematical thinking as the basis for instructional decision-making. Both are important dimensions of effective teaching; they are

(2026). In N. Berger, M. Thompson, J. Ley & C. Wilsmore (Eds.), *Beyond boundaries: Strengthening futures in mathematics education. Proceedings of the 48th annual conference of the Mathematics Education Research Group of Australasia, Sydney* (pp. 138–145). MERGA.

not, however, the same construct. This distinction is elaborated in Ersozlu (under review) and is foundational to the measurement purposes of the present instrument.

Despite substantial theoretical and empirical interest in responsiveness, measuring responsive teaching in classroom settings remains challenging. Widely used observation frameworks have contributed important ways of studying instructional quality, including the Mathematical Quality of Instruction framework and the Teaching for Robust Understanding framework, both of which foreground the intellectual quality of classroom mathematics and students' opportunities to engage meaningfully with mathematical ideas (Hill et al., 2008; Schoenfeld, 2013). In the Australian context, the Quality Teaching framework (NSW Department of Education and Training, 2003; Gore & Bowe, 2017) represents an established approach to observation-based teaching quality research, though it was similarly not designed to isolate responsiveness to students' mathematical thinking specifically. However, these frameworks were not designed specifically to isolate how teachers elicit, interpret, and build on students' mathematical thinking moment by moment. Their primary concern is with the mathematical richness of instructional tasks and explanations rather than with the interactional quality of teacher responses to student ideas as they emerge during a lesson (Hill et al., 2008; Schoenfeld, 2013). Therefore, a teacher could score reasonably well on general instructional quality measures while rarely probing student reasoning, rarely building on student contributions, and rarely adjusting instruction in response to what students actually say or do. A scale designed specifically to capture responsiveness is therefore needed to make these distinctions visible. Furthermore, achieving reliable scores from established observation instruments requires well-developed scoring designs, extensive rater training, and often video-based coding procedures, creating resource demands that make such instruments difficult to deploy in school-based research or teacher professional learning contexts (Hill et al., 2012; Worthington & Whittaker, 2006). There is a practical need for instruments that experienced teachers can use in real time during live classroom observation, with sufficient structure to support reliable ratings and sufficient accessibility to be usable outside specialist research settings. As a result, there remains a need for observation instruments that more directly operationalise responsive mathematics teaching as an observable set of classroom practices, and that are designed with practical feasibility as well as conceptual precision in mind which requires careful pilot work. Responsive teaching, as conceptualised here, is applicable across classroom contexts in which teachers engage with student reasoning during whole-class instruction, small-group work, and individual conferencing. The present study focused specifically on Year 6 mixed-ability primary classrooms, contexts in which teachers must navigate diverse student thinking across multiple instructional phases.

The aim of this study was to develop and pilot a classroom observation instrument for capturing responsive mathematics teaching practices. The quality of instructional responsiveness has implications beyond classroom discourse; classroom conditions characterised by limited mathematical engagement and evaluative feedback have been identified in the literature as antecedents of mathematics anxiety in primary students (Ersozlu & Karakus, 2019; Ersozlu, 2024). The research questions examined: (1) how responsive mathematics teaching can be operationalised in a classroom observation scale, (2) what underlying dimensions emerge from pilot observation data, and (3) how those data can inform refinement of the scale for future research and classroom analysis.

Methodology

Research Design

This study employed a pilot instrument development design in which a responsive teaching observation scale was trialled in authentic classroom settings and subsequently refined based on observation evidence and statistical analysis. The study combined quantitative analysis of observation ratings with qualitative insights from observer field notes and moderation discussions.

Context, Participants and Procedure

The participants were two Year 6 mathematics classrooms teachers within the same Catholic primary school in term three. The two classrooms comprised 21 and 20 students respectively and were mixed-ability classes, representing a typical cross-section of Stage 3 learners. Each classroom had a regular classroom teacher, with one classroom also observed under casual teachers on five of the twelve observation sessions. We conducted observations twice per week over six weeks, yielding a total of 24 classroom observations across the two classrooms (12 observations per classroom). The first author also conducted a subset of classroom observations and facilitated the weekly moderation meetings with the two observers. Each observation session ran for approximately one hour, generally from 11:20 am to 12:30 pm, after the students' first morning break. Observed lessons spanned several mathematical topics across the six weeks, including area of two-dimensional and composite shapes, decimal operations, unit conversions, fractions, and percentage calculations. Lessons generally followed a four-phase structure: an orientation phase in which the teacher introduced the learning intention and activated prior knowledge; an explicit teaching phase involving modelling and demonstration; an exploration phase in which students worked independently or in small groups; and a consolidation phase in which answers were reviewed. Observer field notes were recorded for each lesson in a structured format addressing the orientation, explicit teaching, exploration, and consolidation phases of lessons, as well as observations of responsive pedagogy practices including prioritising student ideas, probing student thinking, and providing feedback. We had the opportunity to review and compare observation notes, discuss scoring decisions and interpretation of scale items, and identify issues requiring instrument refinement in these meetings.

Observation Instrument

A pilot Responsive Mathematics Teaching Observation Scale was developed to capture observable features of responsive pedagogy identified in prior research. Item development was informed by scholarship on professional noticing, responsive teaching, mathematically productive classroom discourse, and instructional quality in mathematics classrooms (Jacobs et al., 2010; Sherin et al., 2011; Hill et al., 2008; Schoenfeld, 2013). The initial 13-item version of the instrument represented practices associated with eliciting student thinking, interpreting mathematical ideas, building on student contributions, and supporting classroom participation. Each item was initially rated on a five-point scale (1–5) reflecting the degree to which the practice was evident during the lesson. However, during the pilot, observers consistently raised the neutral midpoint (3) as problematic: in practice, it was difficult to determine whether a rating of 3 indicated that the practice was partially observed or that the observer was genuinely uncertain. Following moderation discussions, the scale was revised to a four-point format 1 = No evidence, 2 = Limited evidence, 3 = Moderate evidence, 4 = Strong evidence) removing the neutral option to require a directional judgement for each item. Our revisions are consistent with recommendations in the scale-development literature that neutral midpoints can reduce score variance and obscure meaningful differences in observed practice (DeVellis & Thorpe, 2021; Worthington & Whittaker, 2006). One additional item (RP5) was included as a global anchor rating, inviting observers to make a holistic overall judgement of the lesson's responsiveness. This item was treated separately from the 12-dimensional items and was not included in the dimensional total score, serving instead as a criterion for convergent validity checking.

Data Analysis

Data analysis proceeded in three stages. First, internal consistency reliability was examined using Cronbach's alpha as an initial indicator of the coherence of the item set (Cronbach, 1951). Second, an exploratory factor analysis using Principal Component Analysis with Varimax rotation was conducted to examine the underlying structure of the scale and explore whether items clustered around interpretable dimensions of responsive teaching. EFA was selected because no prior factor structure existed for this instrument, making it the appropriate approach for emerging scale development (Costello & Osborne, 2005). Given the pilot nature of the study and the limited sample size, these analyses were treated as exploratory and heuristic rather than confirmatory, in line with guidance on early-stage scale development and factor-analytic interpretation (Costello & Osborne, 2005; DeVellis & Thorpe, 2021). Third, item-level statistics were examined alongside observer field notes and moderation discussions to identify items requiring removal, rewording, or conceptual clarification.

Results

Internal Consistency

The pilot observation scale demonstrated strong internal consistency. The 14-item instrument yielded a Cronbach's alpha of .847, indicating good reliability across the observation items. Inspection of item–total correlations indicated that most items showed moderate to strong alignment with the overall construct of responsive mathematics teaching. However, one item relating to general classroom encouragement demonstrated a negative item–total correlation, indicating that it did not align with the core construct of responsiveness to students' mathematical thinking. This item was therefore removed from the refined version of the instrument.

Exploratory Factor Analysis

Prior to factor analysis, the factorability of the data was assessed. Given the small pilot sample, these diagnostics should be interpreted with caution although item intercorrelations and the pattern of communalities supported the suitability of the data for factor extraction. We used Exploratory Factor Analysis (EFA) using Principal Component Analysis with Varimax rotation was then conducted to examine the underlying structure of the observation instrument. Four components with eigenvalues greater than one were extracted, together explaining 87.6% of the total variance in the observation data. Inspection of the rotated component matrix suggested that the items clustered around several meaningful dimensions reflecting key aspects of responsive teaching practice. These dimensions broadly reflected practices associated with *Attending to Student Thinking*, *Interpreting Student Ideas Mathematically*, *Taking up and Pursuing Student Thinking*, and *Structuring Classroom Participation and Feedback*. These dimensions are consistent with key features of responsive pedagogy described in the literature, particularly work on teacher noticing and instructional responsiveness to students' mathematical thinking (Jacobs et al., 2010; Sherin et al., 2011; Robertson et al., 2016; Sweller, 2024). Factor loadings are presented in Table 1. These findings address RQ2 by identifying the underlying dimensional structure of responsive teaching practices as they emerged from pilot classroom observation data.

Table 1

Exploratory Factor Loadings for the Responsive Mathematics Teaching Observation Scale

Dimension	Item	Factor 1	Factor 2	Factor 3	Factor 4
Taking up student thinking	RP3.3: Teacher encourages students to elaborate on or clarify their thinking.	.853	.108	.004	−.135
Supporting participation	RP4.3: Teacher encourages students to respond to or build on one another's ideas.	.823	.327	.125	−.258

Supporting participation	RP4.4: Teacher structures opportunities for multiple students to contribute ideas	.809	.100	.403	.315
Supporting participation	RP4.1: Teacher provides feedback that helps students refine or extend their mathematical thinking.	.763	.158	−.029	.552
Overall responsiveness	RP5: Overall, instruction in this lesson was shaped by and responsive to students' mathematical thinking.	.662	.637	.173	.067
Attending to student thinking	RP1.1: Teacher invites students to explain their strategies or reasoning.	.420	.867	.062	.194
Attending to student thinking	RP1.3: Teacher makes student thinking visible to the class (e.g., revoicing or recording ideas).	.350	.823	.362	.174
Interpreting mathematical ideas	RP2.1: Teacher connects student ideas to important mathematical concepts, relationships, or principles.	−.274	.809	.208	−.297
Taking up student thinking	RP3.2: Teacher asks follow-up questions that probe students' reasoning.	.483	.738	.146	−.038
Interpreting mathematical ideas	RP2.2: Teacher highlights important mathematical relationships or patterns emerging from student ideas.	−.058	.131	.837	−.251
Taking up student thinking	RP3.4: Teacher facilitates discussion of different student strategies or solutions.	.509	.319	.737	−.024
Attending to student thinking	RP1.2: Teacher attends to and responds to the mathematical content of students' ideas, not just whether their answers are correct.	.133	.495	.707	.090
Supporting participation	RP4.2 (this item is removed)	.026	−.084	−.021	.991
Taking up student thinking	RP3.1: Teacher builds on student ideas or strategies to guide the next step of instruction.	.198	−.207	.542	−.714

Note. Factor loadings above .40 were considered meaningful. Item communalities ranged from .75 to .99, indicating that the extracted components explained a large proportion of item variance.

Scale Refinement

Based on the pilot analyses, the observation instrument was refined in four ways. First, one item relating to general classroom encouragement (RP4.2: "Teacher provides positive reinforcement and encouragement") was removed due to a negative item–total correlation (r = −.109) and independent loading on a separate factor. Because this item reflected general classroom climate rather than responsiveness to student mathematical thinking, its inclusion weakened the conceptual coherence of the scale. Although encouragement and positive reinforcement are important features of classroom climate, they do not necessarily indicate that a teacher is responding to the substance of students' mathematical reasoning. The refined instrument therefore distinguishes general supportiveness from mathematical responsiveness, which is more specifically concerned with how teachers notice, interpret, and build on student thinking during instruction (Hill et al., 2008; Robertson et al., 2016).

Second, two items were reworded to improve clarity and ensure that they captured observable instructional actions associated with building on and interpreting student ideas. Item RP3.1 was revised from "Teacher builds on student ideas to advance the lesson" to "Teacher builds on student ideas or strategies to guide the next step of instruction," to more precisely emphasise the observable instructional action of using student contributions to shape subsequent moves. Item RP2.1 was revised from "Teacher connects student ideas to relevant mathematical concepts and principles" to "Teacher connects student ideas to important mathematical concepts, relationships, or principles," to better align the item wording with the interpretive dimension of responsive teaching in the literature (Jacobs et al., 2010; Robertson et al., 2016). Third, the remaining 12 items demonstrated satisfactory statistical performance (Cronbach's alpha of .868) and conceptual alignment and were retained without change in the refined scale. Fourth, the response scale was revised from five points to four points by removing the neutral midpoint. Observers raised this issue consistently during moderation meetings, noting that a rating of 3 made it difficult to distinguish between partial observation and genuine uncertainty about whether the practice had occurred. Removing the neutral option requires observers to make a directional judgement for each item, improving the discriminant utility of the scale.

Observer field notes further supported the quantitative patterns identified in the ratings. Across both classrooms, notes consistently recorded that teachers acknowledged student responses but rarely pursued the reasoning behind them. One observer noted that a student used a calculator during a lesson and discovered the teacher's solution on the board was incorrect, yet the teacher responded only with "Oh!" and moved on without using the moment as an opportunity for mathematical discussion or collective reasoning. In other lessons, teachers asked students to explain a step in a procedure but redirected or completed the explanation themselves rather than allowing the student's reasoning to be fully explored. Teachers were also noted to provide predominantly generic feedback phrases such as "good job" and "yes, you got it" rather than feedback tied to the specific mathematical substance of student contributions. These patterns were consistent across the two classrooms and corroborated the quantitative finding that items relating to probing, extending, and interpreting student thinking were among the lowest-rated dimensions of the scale.

Discussion

Interpreting the Structure of Responsive Mathematics Teaching

The exploratory factor analysis suggested that responsive teaching practices observed in the pilot lessons could be organised into several interrelated dimensions. These dimensions broadly reflected attending to student thinking, interpreting students' ideas mathematically, taking up and extending student thinking, and structuring participation in ways that made reasoning visible. Together, these dimensions align with prior theory that conceptualises responsive teaching as more than eliciting participation; it involves attending, interpreting, and responding in ways that advance mathematical sense-making (Jacobs et al., 2010; Sherin et al., 2011; Robertson et al., 2016). The first factor, attending to student thinking, resonates with research on teacher noticing, which argues that instruction becomes more responsive when teachers focus on the substance of students' strategies rather than only on answer correctness (Jacobs et al., 2010). The factor concerning interpretation of student ideas reflects the second component of noticing and making sense of what students' contributions reveal about their understandings. The factor involving instructional uptake or extension of student thinking connects with responsive teaching scholarship that emphasises using students' ideas to shape subsequent instructional moves (Robertson et al., 2016). Finally, the factor related to participation and feedback reflects the role of discourse structures in making students' mathematical reasoning available for collective examination, comparison, and refinement (Kazemi & Stipek, 2001; Chapin et al., 2009). The pilot findings therefore provide initial empirical support for conceptualising responsive mathematics teaching as a set of observable classroom practices, while acknowledging that the factor structure should be treated as provisional given the small pilot sample.

Insights from Classroom Observations and Implications for Mathematics Classroom

The classroom observations help situate the scale within known tensions in mathematics teaching. Across both classrooms, lessons typically followed a structured routine in which students worked independently on differentiated tasks while the teacher provided guided support to a small group of three to five students on the floor. Within this structure, observer field notes recorded that teachers frequently acknowledged student responses during whole-class interactions but rarely pressed for explanation or reasoning. In one classroom, the consolidation phase was observed to be consistently abbreviated or absent entirely, meaning that students had little structured opportunity to compare strategies, reflect on their learning, or hear peers' mathematical reasoning. In the other classroom, consolidation was more regularly observed but tended to involve reading out answers from a displayed PDF rather than facilitating substantive mathematical discussion. This pattern reflects a distinction noted in the literature between participation that is answer-oriented and discourse that is reasoning-oriented. Studies of mathematics classrooms have shown that conceptual engagement is more likely when teachers press for explanation, justification, and the comparison of strategies rather than moving quickly to confirmation or demonstration (Kazemi & Stipek, 2001; Chapin et al., 2009; Lampert, 2001).

Although teachers frequently acknowledged student responses, the pilot observations suggested that acknowledgement alone did not consistently lead to instructional uptake of students' mathematical reasoning, a distinction highlighted in prior research on responsive and discussion-based mathematics teaching (Kazemi & Stipek, 2001; Robertson et al., 2016). Observer notes indicated that during explicit teaching phases, questions directed at students were predominantly of a recall or procedural nature, following an initiate–respond–evaluate (IRE) pattern in which correct answers were accepted and the lesson moved on. During exploration phases, teacher-student interactions in the guided group were more substantive, with observers noting moments where students were encouraged to explain their strategies or check their reasoning; however, these interactions were confined to the small guided group and did not extend to the broader class. Observer notes across both classrooms also documented high rates of student disengagement during independent work, with students frequently off task, and several instances where the pace of lessons was driven by time pressure rather than student understanding. The RMTOS was developed within a broader research program concerned with the relationship between teaching quality and students' mathematical experience; these observed conditions provide a meaningful context for understanding why measuring instructional responsiveness matters beyond classroom discourse research alone.

Limitations, Future Research and Conclusion

Several limitations should be acknowledged. First, the pilot dataset consisted of 24 observations across two classrooms observed over six weeks, and the factor analysis results should therefore be interpreted with caution. For that reason, the present study should be understood as identifying a promising conceptual and empirical structure rather than establishing a final validated instrument. Second, although the observers were experienced primary teachers and participated in moderation discussions to support consistent interpretation of the scale items, the observations were conducted within a single school context across a limited range of Stage 3 content topics including decimals, fractions, area, and order of operations. Future studies should examine the applicability of the instrument across a wider range of mathematical topics, year levels, and classroom contexts. Further research with larger samples, shared video-coded lessons, and multiple raters will be necessary to test the stability of the factor structure, strengthen inter-rater reliability evidence, and examine whether the scale predicts important outcomes such as the quality of classroom discourse or student learning opportunities. Incorporating qualitative methods such as teacher interviews or post-lesson conversations would also allow future research to investigate the reasons behind observed instructional patterns, including teachers' reported challenges in pursuing individual students' thinking during whole-class instruction. Examining the relationship between RMTOS ratings and established frameworks for instructional quality would provide important evidence of convergent validity and situate the instrument within the broader landscape of observation-based research on teaching quality in Australian schools.

Acknowledgments

This research was supported by internal funding from the University of Newcastle under the project *"Decoding Mathematics Anxiety: Pedagogical Dynamics". The responsive teaching observation work*

reported here represents a distinct contribution being extended in subsequent validation studies. Ethics approval for this study was granted by the University of Newcastle College Human Ethics Advisory Panel (Approval No. H-2024-0153), and all participants provided informed consent prior to participation. Approval to conduct the research in the relevant Catholic education setting was also obtained from the Catholic Schools Office, Diocese of Maitland-Newcastle.

References

Chapin, S. H., O'Connor, C., & Anderson, N. C. (2009). *Classroom discussions: Using math talk to help students learn, Grades K–6* (2nd ed.). Math Solutions.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1–9. https://doi.org/10.7275/jyj1-4868

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). Sage.

Ersozlu, Z. (2024). The role of technology in reducing mathematics anxiety in primary school students. *Contemporary Educational Technology, 16*(3), ep517. https://doi.org/10.30935/cedtech/14717

Ersozlu, Z. & Karakus, M. (2019). Mathematics anxiety: Mapping the literature by bibliometric analysis. *Eurasia Journal of Mathematics, Science and Technology Education, 15*(2), em1673. https://doi.org/10.29333/ejmste/102441

Ersozlu, Z. (under review). Responsive mathematics teaching: Conceptual distinctions, measurement challenges, and a framework for the Australian context. Manuscript submitted for publication.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction, 26*(4), 430–511. https://doi.org/10.1080/07370000802177235

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. https://doi.org/10.3102/0013189X12437203

Gay, G. (2010). *Culturally responsive teaching: Theory, research, and practice* (2nd ed.). Teachers College Press.

Gore, J. M., & Bowe, J. M. (2017). Reassembling teacher professional development: The case for Quality Teaching Rounds. *Teachers and Teaching: Theory and Practice, 23*(3), 352–366. https://doi.org/10.1080/13540602.2016.1206522

Jacobs, V. R., Lamb, L. L. C., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education, 41*(2), 169–202. https://doi.org/10.5951/jresematheduc.41.2.0169

Kazemi, E., & Stipek, D. (2001). Promoting conceptual thinking in four upper-elementary mathematics classrooms. *The Elementary School Journal, 102*(1), 59–80. https://doi.org/10.1086/499693

Ladson-Billings, G. (1995). Toward a theory of culturally relevant pedagogy. *American Educational Research Journal, 32*(3), 465–491. https://doi.org/10.3102/00028312032003465

Lampert, M. (2001). *Teaching problems and the problems of teaching.* Yale University Press.

NSW Department of Education and Training. (2003). *Quality teaching in NSW public schools: A classroom practice guide.* NSW Department of Education and Training, Professional Support and Curriculum Directorate.

Robertson, A. D., Scherr, R. E., & Hammer, D. (Eds.). (2016). *Responsive teaching in science and mathematics.* Routledge.

Schoenfeld, A. H. (2013). Classroom observations in theory and practice. *ZDM Mathematics Education, 45*(4), 607–621. https://doi.org/10.1007/s11858-012-0483-1

Sherin, M. G., Jacobs, V. R., & Philipp, R. A. (Eds.). (2011). *Mathematics teacher noticing: Seeing through teachers' eyes.* Routledge.

Sweller, J. (2024). Cognitive load theory and individual differences. *Learning and Individual Differences, 111*, 102434.

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*(6), 806–838. https://doi.org/10.1177/0011000006288127

van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology, 36*(3), 212–218.