INTERDISCIPLINARY JOURNAL OF SOCIAL SCIENCE AND EDUCATION

ISSN 3032-7563
Volume 4, Issue 3, 2026, pp. 245–256
DOI: https://doi.org/10.53639/ijsse.v4i3.174

Teacher Mentoring Strategies to Improve the Quality of Literacy and Numeracy Items at SD Negeri 2 Sirau, Karangmoncol, Purbalingga

Kholidin[1]*

[1] UIN Profesor Kiai Haji Saifuddin Zuhri Purwokerto, Purwokerto, Indonesia.

* Corresponding author: 254120500010@mhs.uinsaizu.ac.id

ARTICLE INFO

Article history

Received : 2026-09-08
Accepted : 2026-11-05
Published: 2026-11-25

Keywords

Teacher Mentoring;
Literacy and Numeracy Items;
Item Quality;
Learning Community;
Primary Education.

ABSTRACT

Low literacy and numeracy attainment in Indonesia makes the quality of teacher-made items a natural place to start improvement, yet primary teachers rarely receive sustained mentoring in writing stimulus-based items that demand reasoning. This study examined how teacher mentoring strategies can improve the quality of literacy and numeracy items at SD Negeri 2 Sirau. A descriptive qualitative case study supported by simple quantitative data followed one mentoring cycle of four meetings. Data came from non-participant observation, semi-structured interviews with two teachers and the principal, and documentation including the module, draft items, and an eight-aspect item-review rubric, and were analyzed thematically with technique and source triangulation. The mentoring proceeded systematically through concept reinforcement, group practice in writing items, rubric-based peer review, and reflection. Teacher engagement moved from listening to active questioning, discussion, and mutual correction as practice and feedback increased. Revised items used contextual stimuli, demanded reasoning, and took more varied formats, whereas difficulties with reasoning items, distractors, and scoring guidelines for essay items persisted. The findings suggest that practice-based mentoring combined with peer feedback and principal support is a workable model for improving item quality in primary schools. The study is limited by its small number of informants and its single cycle, so future research should involve more schools and analyze students' responses to the items.

This is an open-access article under the CC-BY-SA license.

How to cite: Kholidin. (2026). Teacher Mentoring Strategies to Improve the Quality of Literacy and Numeracy Items at SD Negeri 2 Sirau, Karangmoncol, Purbalingga. *Interdisciplinary Journal of Social Science and Education, 4*(3), 245–256. https://doi.org/10.53639/ijsse.v4i3.174

1. Introduction

Teacher-made test items decide whether efforts to strengthen literacy and numeracy are ever measured in a classroom. Items that test only recall say little about how well students reason with texts and numbers. Teachers' own assessments are also the feedback that students and teachers see most often. Indonesia's PISA 2022 means were 366 in mathematics and 359 in reading, and just 18% and 25% of students reached Level 2 or higher (OECD, 2023). In the national Minimum Competency Assessment (Tes Kemampuan Akademik, TKA), only 46.67% of students reached above-minimum numeracy (Hardhita et al., 2024). In the 2026 Academic Ability Test (Tes Kemampuan Akademik, TKA), primary students averaged 60.14 in Indonesian language and 43.41 in mathematics (Gebrak.id, 2026). Researchers have looked at students' difficulties with TKA numeracy items (Anggraini & Setianingsih, 2022; Lestari et al., 2023), and (Hijjayati et al., 2022) and (Ardhita et al., 2024) studied why reading and writing literacy stays low in primary schools. Improving literacy and numeracy also depends on whether teachers can write good items.

Item-writing training has been studied often, but little is known about how mentoring works inside one school and what it does to item quality. Community-service reports describe teachers who could draft blueprints and items after training, including TKA-based numeracy items and higher-order thinking skills (HOTS) items (Hardhita et al., 2024; Nurwahidah et al., 2023; Shamdas et al., 2023), and training has also targeted teachers' own literacy and numeracy competence (Marlena et al., 2022; Pulungan & Elfrianto, 2024). Other studies find that primary teachers still struggle with HOTS items and misunderstand what they are (Isrokatun et al., 2023; Sinta et al., 2022). In Tanzanian primary schools, 73.2% of teacher-made multiple-choice questions tested only lower cognitive levels (Kyaruzi & Kisamo, 2025), and teacher-made tests in Ghana contained mostly non-functioning distractors (Butakor, 2022). Work on sustained mentoring through learning communities mostly concerns teachers' general competence (Harlita & Ramadan, 2024; Kalman et al., 2024). Few studies in the literature reviewed follow mentoring meeting by meeting and also record how literacy and numeracy items changed in a single primary school. That is the gap this study addresses.

This article asks how teacher mentoring can improve the quality of literacy and numeracy items at SD Negeri 2 Sirau. Because the gap concerns process as well as outcome, the study follows one mentoring cycle closely and does not rely only on a before-and-after comparison. What it adds to training-based reports is a link between what the mentor and teachers did in each meeting and what changed in the items. The first objective is to describe the mentoring over four meetings: its stages, methods, and the mentor's role. The second is to analyze teacher engagement and the factors that helped or hindered the work. The third is to examine how item quality changed from before to after mentoring and what follow-up it calls for. The findings below are organized around these three objectives.

The study argues that mentoring that combines concept reinforcement, hands-on practice, peer review, and reflection will improve teachers' literacy and numeracy items. Knowing the concepts rarely changes how teachers write items; changing practice takes chances to rehearse and specific feedback (Glover et al., 2023). Collegial relationships inside a learning community also support engagement and reflection (Arifin & Hanif, 2024; Nugent et al., 2023). Three propositions follow and can be checked against the data: the mentoring was systematic, teacher engagement rose with practice and feedback, and revised items improved in stimulus, cognitive level, and format variety. The contribution is empirical and local: a documented mentoring sequence for item writing in a school with limited resources, plus a rubric for reading change in items. School leaders and district mentors who plan support for teachers as TKA-style items become routine can use it. If the propositions hold, practice-based mentoring is a credible way to improve item quality in primary schools.

2. Methods

The unit of analysis is SD Negeri 2 Sirau, a public primary school where one cycle of mentoring on writing literacy and numeracy items took place. One school made it possible to follow the mentoring in detail from meeting to meeting, which multi-school designs with short observation windows rarely allow. The school also suited the question: most teachers had no specific training in writing TKA-style items, and the principal took part in the activity. Sessions ran in the teachers' room and moved to a larger Grade V classroom for the third meeting. Participants were classroom teachers, and the principal attended the first and fourth meetings. Meetings were held after school on 22, 23, 24, and 26 October 2026. The case is therefore bounded to a single four-meeting cycle in one public school.

The study is a descriptive qualitative case study with simple quantitative data added. A qualitative approach suits the aim of understanding how mentoring was carried out, experienced, and given meaning by teachers in their own setting. Observation scores and item-review rubric scores were used only to show the direction of change, and no hypothesis was tested statistically. All data are primary and were collected by the researcher in the field through observation, interviews, and documentation. Reading narrative and numeric evidence together resembles the concurrent triangulation used in studies of teacher-made items in primary schools (Kyaruzi & Kisamo, 2025). The mentoring itself followed one cycle of planning, implementation, observation, and reflection. This design fits objectives that need both a description of process and a picture of change in outcomes.

Data came from three informants and from activity documents. Informants were chosen purposively because they took part in the mentoring or were responsible for school policy. Comparing teachers with the principal also allowed cross-checking. G1 teaches Grade V, has 12 years of experience, and had attended one cluster-level assessment briefing. G2 teaches Grade IV, has six years of experience, and has had no specific training. The third informant, the principal (KS), had served for three years and was responsible for school policy, including time and facilities for teachers. Documents included the mentoring module, teachers' items before and after mentoring, observation sheets, and field notes. A single case limits what these sources can show, but together they cover the questions asked.

Three techniques were used: non-participant observation, semi-structured interviews, and documentation. Observation let the researcher follow the mentoring without shaping its content, and interviews asked informants about their experience and views. Documentation supplied written evidence of the items and the activities, so what informants said could be matched against what teachers produced. All four meetings were observed with a sheet rating mentoring implementation and teacher engagement, with field notes and documentation recorded on a separate form. Interviews took place on 24 October 2026 in the teachers' room, 30 minutes each with G1 and G2 and 15 minutes with the principal, on conceptual understanding, item-writing practice, obstacles, the mentoring process and its effects, and expectations. Teachers' items before and after mentoring were reviewed with an eight-aspect rubric scored from 1 to 4, for a maximum of 32. Permission came from the Faculty of Tarbiyah and the school, informants consented, and teachers appear only as G1, G2, and KS.

Analysis was thematic: reading the data, coding, building themes, and drawing conclusions, followed by triangulation. Thematic analysis condenses narrative data while keeping each theme traceable to its source and meeting. Coding mixed deductive codes taken from the research framework with inductive codes that came from the data (Naeem et al., 2023). Observation notes and interview transcripts were coded and grouped under mentoring implementation, teacher engagement, change in item quality, and supporting and hindering factors. Observation and rubric scores were turned into percentages of the maximum and categorized as the instrument sheets prescribe. To check trustworthiness, observation findings,

informants' statements, and item documents were compared, including whether what teachers said matched what they did. The aim was findings that hold across sources.

3. Literature Review

Teacher mentoring is continuing professional support in which a mentor or colleagues help a teacher improve practice. It assumes that teachers learn best when development is built into daily work and not delivered as a single training day. The international literature usually calls it instructional coaching and stresses observation, feedback, and reflective discussion (Nugent et al., 2023). (Glover et al., 2023) asked which coaching actions predict teachers' classroom practice and student achievement, and opportunities to practice were a meaningful predictor. In Indonesia, strategies for raising teachers' professional competence are discussed as part of quality improvement (Ilyas, 2022), and managing school learning-community programs has been linked to teachers' pedagogical competence (Arifin & Hanif, 2024). For assessment, the aim is assessment literacy, the knowledge and skills teachers need to run classroom assessment (Pastore, 2023). In this study, mentoring means continuing professional development built around the practice of writing items.

Mentoring has activity elements and relational elements. The activity elements are observation, feedback, reflective discussion, and planning follow-up, and the relational element is the quality of the mentor–teacher relationship (Nugent et al., 2023). Mentoring can be delivered by an expert mentor or through collective learning among colleagues in a learning community. (Nugent et al., 2023) found that relationship quality predicted teachers' engagement and reflection, while the quality of coaching strategies predicted overall instructional quality. Learning communities offer a collective setting for this work and are associated with better teacher competence in implementing the Merdeka Curriculum (Giyanto et al., 2023; Kalman et al., 2024). They have been reported in driving schools (sekolah penggerak) and discussed for their role in developing teacher competence in primary schools (Harlita & Ramadan, 2024; Isnawan et al., 2024). This study therefore sorts mentoring into activity elements (concepts, practice, review, reflection) and collegial relationships that shape how teachers receive feedback.

Literacy and numeracy items ask students to understand and use information from texts, data, or real-life contexts in order to reason and solve problems. PISA defines mathematical literacy in a similar way, as the capacity to formulate, employ, and interpret mathematics in varied contexts (OECD, 2023). In Indonesia the idea is carried by the AKM, which treats reading literacy and numeracy as basic competencies, and since 2026 by the TKA for primary students. Training in writing AKM-based numeracy items for primary teachers aims to help them draft blueprints and items aligned with learning outcomes (Hardhita et al., 2024). HOTS-based literacy and numeracy items for primary students have been developed to demand reasoning (Kusuma & Nurmawanti, 2023), and strengthening teachers' competence in writing such instruments has been framed as support for AKM (Ginting et al., 2024). In this study, literacy and numeracy items are items with a contextual stimulus that call for application and reasoning, not recall.

Such items can be described by five elements: stimulus, context, cognitive level, item format, and answer key with scoring guideline. The stimulus is the text or data students think about, and the context ties the item to their lives. Cognitive level runs from recalling to applying and reasoning. Training for primary teachers has stressed well-worded, readable item stems and the rules for writing HOTS items in essay and multiple-choice formats (Shamdas et al., 2023). Teachers' difficulties are reported mainly in designing items that match learning outcomes and rest on higher-order thinking (Hardhita et al., 2024), and similar difficulties appear with HOTS evaluation items in primary mathematics (Sinta et al., 2022). These studies

concentrate on alignment and cognitive level, so the present study also looks at format variety and scoring guidelines. All five elements feed into the review rubric used here.

Item quality is how accurately, clearly, and functionally an item measures the competence it targets. A good multiple-choice item has suitable difficulty, distractors that work, and no writing flaws that invite guessing (Butakor, 2022; Kyaruzi & Kisamo, 2025). Quality also depends on the teacher's assessment competence, meaning the knowledge and skills for carrying out classroom assessment (Pastore, 2023). (Butakor, 2022) found that in teacher-made mathematics tests in Ghana, 30.8% of options were functioning distractors and 69.2% were not. (Kyaruzi & Kisamo, 2025) reported that among the writing flaws found in primary teachers' items, 69.2% were test-wiseness flaws and 30.8% were irrelevant difficulty. Evaluation in a broader sense also covers cognitive, affective, and psychomotor aspects, as studied in Islamic religious education (Ujiyanti & Hanif, 2025). Item quality is therefore treated here as multidimensional and judged with a review rubric.

The study reviews items on eight aspects: alignment with indicators, contextual stimulus, cognitive level, literacy or numeracy content, clarity of language and concepts, answer options and distractors, answer key and scoring guideline, and format variety. Together they cover content, construction, and language, the usual dimensions of item review. Reviewing aspect by aspect shows where improvement is most needed. Teachers' multiple-choice items in Tanzania often assess low cognitive levels and contain construction flaws (Kyaruzi & Kisamo, 2025), and distractor weaknesses appear in Ghana (Butakor, 2022). Research on writing in primary schools favors a student-centered process approach that builds students' thinking and creativity (Kusrini & Hanif, 2024), which supports items that demand thinking. Mentoring research explains how teachers change, and item-quality research explains what needs to change; few studies connect the two in one school. The eight aspects are the lens for the findings that follow.

4. Result and Discussion

4.1 How the Mentoring Unfolded

Mentoring at SD Negeri 2 Sirau was systematic: four meetings, each with its own focus and set of activities. The sequence went from needs mapping to practice, then review, then reflection and follow-up. A module and sample items supported each stage and gave teachers a shared reference. At the first meeting the mentor mapped needs and reinforced concepts by comparing two sample items in the AKM/TKA format with end-of-term items teachers had brought. The second and third meetings covered writing stimuli and items and reviewing them with feedback. The fourth was for revision, reflection, and follow-up. The observation sheet gave mentoring implementation 70.0% at the first meeting and 92.5% at the fourth (Table 1). The mentoring ran in order, as planned for a single cycle.

Table 1

Observation scores for mentoring implementation and teacher engagement

Meeting	Mentoring implementation (max. 40)	%	Teacher engagement (max. 28)	%
1	28	70.0	17	60.7
2	32	80.0	21	75.0
3	35	87.5	24	85.7
4	37	92.5	26	92.9

Note. Mentoring implementation rose from the "good" category (70.0%) to "very good" (92.5%), and teacher engagement from "fair" (60.7%) to "very good" (92.9%).

The strategy had four components: concept reinforcement, group practice, rubric-based peer review, and reflection. Each did a different job. Concepts gave teachers a knowledge base, practice gave them experience, review gave them feedback, and reflection helped hold the changes. The mentor moved between explaining, facilitating, and asking probing questions. When asked what literacy and numeracy meant to them, most teachers answered "reading" and "counting" before the mentor introduced stimulus, context, and cognitive level. In practice, three mixed groups chose contexts close to students' lives, such as the market, the garden, or the boat timetable. In review, teachers took turns reading colleagues' items and wrote rubric-based comments, for example on illogical distractors or answer keys with no scoring guideline. Every component put teachers in direct contact with items.

The school supported the mentoring in policy but was short of facilities and time. The principal gave the activity legitimacy and space, while limited equipment shaped how sample items were shown. The principal opened the activity, provided a room and schedule, facilitated the learning community, printed the module, and arranged teaching hours so teachers could attend (KS interview). The teachers' room was medium-sized, with a long table where teachers sat facing one another. Without a projector, sample items were shown in turn from printed sheets and a laptop screen. The principal also said the school urged teachers to avoid items that only test memorization, but no binding rule existed. The mentoring therefore rested on commitment from the school and not yet on a formal rule.

If mentoring is systematic and built from four ordered components, the structure is part of why it works. Concept, practice, review, and reflection give teachers a staged path, so each meeting asks for a manageable amount of new thinking. The order also creates clear points for feedback, which the coaching literature treats as central (Nugent et al., 2023). Spending the first meeting on concepts and the later ones on practice fits the finding that practice opportunities predict change in teachers' practice (Glover et al., 2023). Earlier training for primary teachers also combined instruction, group practice, and evaluation (Hardhita et al., 2024; Shamdas et al., 2023). What differs here is that peer review and reflection were core components and not closing formalities. A staged design with practice and feedback is a sound basic frame for similar mentoring.

The proposition that the mentoring was systematic holds in this case, and the school's readiness explains why. The principal's room, schedule, teaching-hour arrangements, and module removed technical obstacles that often derail teachers' activities. This agrees with the view that managing a school learning-community program relates to better pedagogical competence (Arifin & Hanif, 2024), and a study in public primary schools reported improved teacher competence when the activity ran at school level (Kalman et al., 2024). The principal's remark that no binding rule exists, however, suggests the systematic character still depends on leadership commitment. One cycle in one school cannot show whether it would last. The first proposition is supported for one cycle, and institutionalization would be needed for it to endure.

4.2 Teacher Engagement and the Conditions Behind It

Teacher engagement rose gradually, from listening at the early meetings to asking, discussing, and correcting one another at the last. It showed in what teachers did and in engagement scores that climbed from meeting to meeting. Teachers who were quiet at first became bolder about giving their views. Only a few teachers asked questions at the first meeting. By the second, group discussion was lively, and teachers asked one another about reading difficulty and suitable numbers. At the third and fourth meetings they accepted criticism openly, fixed items on the spot, and reflected on how their way of writing items had changed. The engagement score rose from 60.7% to 92.9%, with the biggest jump

between the first and second meetings, when hands-on practice began (Table 1). Engagement grew as chances to practice and interact grew.

Hands-on practice and personal feedback seemed the factors most closely tied to higher engagement and better items. Writing items and then reviewing them together gave teachers concrete experience they could connect to the concepts. Specific feedback showed where the mistakes were, so teachers could fix them at once. G1 said, "What helped most was the practice and the joint item review, so I knew where I had gone wrong." G2 named group discussion and peer feedback, because teachers could see one another's items. Observation showed the clearest change after teachers wrote their own items and reviewed them together at the second and third meetings. At the second meeting some groups struggled to phrase reasoning questions and slipped back to recall items until the mentor showed how to raise an item's level. Practice, feedback, and engagement reinforced one another.

The link between practice, feedback, and engagement held under conditions that both helped and hindered. Supports strengthened it, and obstacles limited its depth and continuity. Knowing both matters if the strategy is to be replicated realistically. Supports included the principal, a communicative mentor, mixed groups of experienced and novice teachers, the module and sample items, and a collegial atmosphere. Obstacles were limited time and administrative burden, uneven starting abilities, no item bank or templates, and limited projection and printing facilities. G1 described writing items late at night in a hurry, with barely adequate results, and G2 admitted low confidence for fear that a stimulus was too hard or a concept wrong. The link works where institutional support exists and is held back by time pressure and missing tools.

If engagement rises with practice and feedback, mentoring works by making teachers actors, not recipients of content. Teachers who take part interpret concepts through their own experience, so understanding becomes meaningful and carries over more easily into practice. Engagement also opens room for reflection, an important part of professional learning. (Nugent et al., 2023) showed that the quality of the mentor–teacher relationship predicts teachers' engagement and reflection and is built through rapport and mutual trust. G1 found the mentor's feedback clear and not patronizing, and G2 described a relaxed atmosphere with freedom to ask, which is consistent with that finding. The principles of active learning discussed by (Humam & Hanif, 2025) for students' critical thinking seem to apply to teachers as learners as well. Mentoring should give teachers an active role from the first meeting.

The proposition that engagement rose with practice and feedback holds descriptively, but the causal mechanism needs careful reading. The largest rise came just as hands-on practice began, yet familiarity among participants and habituation to the observer also changed across meetings. A single cycle with no comparison group cannot separate these influences. That teachers in mixed groups of experienced and novice members questioned one another supports the idea that peer interaction contributed, which agrees with the role of learning communities in developing teacher competence (Harlita & Ramadan, 2024; Isnawan et al., 2024). (Ngafifah & Hanif, 2026) reported that students' understanding varies with their earlier educational background, which suggests that teachers' uneven starting abilities could likewise shape how each responds to mentoring. The second proposition is supported descriptively, and causal claims need a stricter design.

4.3 Changes in Item Quality and Follow-Up Needs

Teachers' items rose from the "poor" category to the "good" category after mentoring. The gains came mainly in the aspects most tied to the character of literacy and numeracy items: stimulus, cognitive level, and format variety. Alignment with indicators

was already good at the start and had little room to rise. The rubric total went from 14 to 25 out of 32, or from 43.8 to 78.1 on a 100-point scale (Table 2). Contextual stimulus, cognitive level, and format variety each rose from 1 to 3, and alignment with indicators from 3 to 4. Review notes say items that once had almost no stimulus now use texts, tables, or data, and that formats now include complex multiple choice, matching, and short answer. Quality changed most in the dimensions closest to literacy and numeracy.

Table 2

Item-review rubric scores before and after mentoring

No.	Aspect	Pre	Post	Note on change
1	Alignment with indicators	3	4	All items measure the indicators accurately
2	Contextual stimulus	1	3	From almost no stimulus to texts, tables, or data
3	Cognitive level	1	3	From mostly recall to application and reasoning
4	Literacy/numeracy content	2	3	Items require understanding and using information
5	Clarity of language and concepts	2	3	Fewer ambiguous sentences
6	Answer options and distractors	2	3	More logical, functioning distractors
7	Answer key and scoring guideline	2	3	Correct keys; scoring guidelines begin to appear
8	Format variety	1	3	Complex multiple choice, matching, and short answer added
	Total (max. 32)	**14**	**25**	**Score 43.8 (poor) to 78.1 (good)**

Note. Each aspect was scored from 1 to 4.

The results imply that follow-up should strengthen teachers in the aspects that are still weak. The changes are early and need consolidation, and some difficulties persist beyond this first cycle. Teachers voiced concrete needs that the school and its mentors can act on. Persistent difficulties are writing items that demand reasoning, building working distractors, and drafting scoring guidelines for essay items. Teachers and the principal agreed to build a shared item bank, hold monthly mentoring through the learning community, and arrange special mentoring on essay items and numeracy items with graphs, and the observation notes added that proficient teachers could act as peer mentors. G2 said, “I want further mentoring to write scoring guidelines for essay items, and practice writing numeracy items with graphs or diagrams.” These steps extend teachers’ learning beyond a single cycle.

Practice-based mentoring inside a learning community offers the school both value and an opening. The value is a real change in teachers’ practice that reaches the items and the teaching. The opening is the principal’s commitment to place the activity in a monthly learning community. G1 gave an example: a fractions item that began as direct computation became a table of fish prices at the Sirau market, asking students to calculate change and choose the cheapest option. G2 replaced the question “Who is the character in the story?” with an information text on litter on the beach, asking students to identify causes and propose solutions. Both teachers reported that students became more active in discussion and reasoning. The principal plans a monthly learning community so that a shared item bank and peer mentoring can grow from what has begun. If institutionalized, the strategy can last.

If item quality rose mainly in stimulus, cognitive level, and format variety, mentoring moved teachers from recall items toward reasoning items. Those aspects mark literacy and numeracy items, so improving them brings items closer to what the AKM and TKA demand. Students also get more chances to practice reasoning, the weakness shown in their international results (OECD, 2023). Before mentoring, the principal said, most teachers wrote items from textbooks, mostly recall, with almost no stimulus-based reasoning items. This resembles findings that most teacher-made primary items in Tanzania assess only low cognitive levels (Kyaruzi & Kisamo, 2025) and that primary teachers struggle to write HOTS items (Sinta et al., 2022). After mentoring, revised items used a fish-price table or a text about litter on the beach, which require students to compute, identify causes, and choose solutions. This change in the kind of item is the most important result of the mentoring.

The proposition that revised items improved in stimulus, cognitive level, and format variety is supported, but the broader claim of overall item quality is not. Distractors and the answer key and scoring guideline still need work, so the improvement is partial. Writing flaws and non-functioning distractors are common problems in teacher-made items (Butakor, 2022; Kyaruzi & Kisamo, 2025), and they persisted here. Rubric scores for distractors and for answer key and scoring guideline rose only from 2 to 3, and the teachers themselves asked for further mentoring on essay items and numeracy items with graphs. The rubric review was also not accompanied by item analysis of student responses, so distractor functioning was never tested empirically. Strengthening teachers' competence in writing literacy and numeracy instruments will need repetition and continuity (Ginting et al., 2024; Pastore, 2023). The third proposition holds for the aspects that define literacy and numeracy items, not yet for every aspect of quality.

5. Conclusion

The central finding is that changes in teachers' items can be traced to specific mentoring components, above all hands-on practice and rubric-based peer review. That link between process and outcome becomes visible only when mentoring is observed meeting by meeting, not judged from an end-of-activity report. Observation paired with interviews shows how the change happened and not merely that it did. Teachers moved from listening to taking part as practice and feedback accumulated, and their items moved from recall toward stimulus-based reasoning, most clearly in contextual stimulus, cognitive level, and format variety. Reasoning items, distractors, and scoring guidelines for essay items stayed difficult after one cycle. Practice-based, peer-reviewed mentoring moves item quality, but one cycle does not finish the job.

The study confirms that practice and feedback matter in mentoring and extends that finding to literacy and numeracy item writing in a primary school. This matters because much of the evidence on mentoring comes from other settings or from general pedagogical competence. Its strength is the combination of process observation, interviews across roles, and rubric-based review in one case. The framework of stimulus, context, and cognitive level, together with the eight-aspect rubric, gave clear categories for answering the research questions, and thematic analysis with inductive and deductive coding (Naeem et al., 2023) kept the findings organized. The results agree with earlier work on mentoring and learning communities (Glover et al., 2023; Kalman et al., 2024) and add detail on a strategy specific to item writing. School leaders and district mentors planning support for teachers can adapt that detail. The contribution is confirmatory and tied to its context.

The study is limited in the number of informants, the scope of the case, its duration, and how items were reviewed. Only two teachers and one principal were interviewed, in one school, over one four-meeting cycle, so the findings cannot be generalized. Items were reviewed with a

rubric, without item analysis of student responses, and the reports of more active students come from teachers' own statements. Observation and rubric scores show the direction of change and come from one observer. Future studies should include more schools and teachers, use comparison groups, and add item analysis of difficulty, discrimination, and distractor functioning from student responses. They should also follow the learning community and item bank for several months and examine effects on students' literacy and numeracy. Until then these findings are best read as a starting point.

References

Anggraini, K. E., & Setianingsih, R. (2022). Analisis kemampuan numerasi siswa SMA dalam menyelesaikan soal Asesmen Kompetensi Minimum (AKM) [Analysis of senior high school students' numeracy ability in solving Minimum Competency Assessment (AKM) items]. *Jurnal Mathedunesa*, *11*(3), 837–849.

Ardhita, Y. A. A. A. A., Rahma, F. M., Sari, K., & Khairunnisa. (2024). Analisis faktor penyebab rendahnya kemampuan literasi baca tulis siswa sekolah dasar [Analysis of the factors behind low reading and writing literacy among primary school students]. *Jurnal Guru Kita*, *8*(3), 306–312. https://doi.org/10.24114/jgk.v8i3.60183

Arifin, J., & Hanif, M. (2024). Manajemen program komunitas belajar sekolah untuk peningkatan kompetensi pedagogik guru [Management of school learning community programs to improve teachers' pedagogical competence]. *Ideguru: Jurnal Karya Ilmiah Guru*, *9*(3), 1421–1432. https://doi.org/10.51169/ideguru.v9i3.1112

Butakor, P. K. (2022). Using classical test and item response theories to evaluate psychometric quality of teacher-made test in Ghana. *European Scientific Journal*, *18*(1).

Gebrak.id. (2026). Nilai rerata TKA SD-SMP 2026 bikin kaget! Bahasa Indonesia tembus 60, matematika masih di angka 40 [2026 national average TKA scores for primary and junior high school: Indonesian language reaches 60, mathematics stays around 40]. https://www.gebrak.id/?p=28196

Ginting, F. W., Unaida, R., & Sakdiah, H. (2024). Penguatan kompetensi guru dalam menyusun instrumen literasi numerasi untuk mendukung program Asesmen Kompetensi Minimum (AKM) [Strengthening teachers' competence in writing literacy and numeracy instruments to support the Minimum Competency Assessment (AKM) program]. *Jurnal Pengabdian Sosial*, *2*(1), 2493–2499.

Giyanto, B., Kurnia, P., Julizar, K., Sari, D. K., & Hartono, D. (2023). Implementasi kebijakan komunitas belajar dalam Kurikulum Merdeka Belajar di Indonesia [Implementation of the learning community policy in the Merdeka Belajar Curriculum in Indonesia]. *Jurnal Pembangunan dan Administrasi Publik*, *5*(2), 37–50.

Glover, T. A., Reddy, L. A., & Crouse, K. (2023). Instructional coaching actions that predict teacher classroom practices and student achievement. *Journal of School Psychology*, *96*, 1–11.

Hardhita, R. S., Untari, S., & Suciptaningsih, O. A. (2024). Pelatihan pembuatan soal numerasi berbasis Asesmen Kompetensi Minimum (AKM) bagi guru SDN Wonotirto 4 [Training in writing Minimum Competency Assessment (AKM)-based numeracy items for teachers of SDN Wonotirto 4]. *Jurnal SOLMA*, *13*(2), 1066–1076. https://doi.org/10.22236/solma.v13i2.15421

Harlita, I., & Ramadan, Z. H. (2024). Peran komunitas belajar di sekolah dasar dalam mengembangkan kompetensi guru [The role of learning communities in primary schools in developing teacher competence]. *Didaktika: Jurnal Kependidikan*, *13*(3).

Hijjayati, Z., Makki, M., & Oktaviyanti, I. (2022). Analisis faktor penyebab rendahnya kemampuan literasi baca-tulis siswa kelas 3 di SDN Sapit [Analysis of the factors behind low reading and writing literacy among third-grade students at SDN Sapit]. *Jurnal Ilmiah Profesi Pendidikan*, *7*(3b), 1435–1443. https://doi.org/10.29303/jipp.v7i3b.774

Humam, M. S., & Hanif, M. (2025). Strategi pembelajaran aktif dalam meningkatkan keterampilan kritikal siswa di era modern [Active learning strategies to improve students' critical skills in the modern era]. *Jurnal Bintang Pendidikan Indonesia*, *3*(3), 89–108.

Ilyas, I. (2022). Strategi peningkatan kompetensi profesional guru [Strategies for improving teachers' professional competence]. *Jurnal Inovasi, Evaluasi dan Pengembangan Pembelajaran (JIEPP)*, *2*(1), 34–40. https://doi.org/10.54371/jiepp.v2i1.158

Isnawan, M. G., Alsulami, N. M., & Sukarma, K. (2024). Pengembangan komunitas belajar di sekolah penggerak Kabupaten Sumbawa [Developing learning communities in driving schools in Sumbawa Regency]. *ALAMTANA: Jurnal Pengabdian Masyarakat UNW Mataram*, *5*(2). https://doi.org/10.51673/jaltn.v5i2.2163

Isrokatun, I., Sunaengsih, C., Maulana, M., Syahid, A. A., Karlina, D. A., & Rohaeti, P. (2023). Miskonsepsi guru sekolah dasar mengenai pemahaman pembelajaran berbasis HOTS [Primary school teachers' misconceptions about understanding HOTS-based learning]. *Dwija Cendekia: Jurnal Riset Pedagogik*, *7*(3).

Kalman, K., Muhammadiah, M., & Hasbi, M. (2024). Implementasi komunitas belajar dalam peningkatan kompetensi guru UPTD Sekolah Dasar Negeri di Kabupaten Mamuju Tengah [Implementation of learning communities in improving the competence of teachers at public primary schools under the UPTD in Central Mamuju Regency]. *Bosowa Journal of Education*, *5*(1), 137–143. https://doi.org/10.35965/bje.v5i1.5278

Kusrini, E., & Hanif, M. (2024). Perkembangan motorik dan kognitif siswa SD Al-Irsyad Al-Islamiyyah Purwokerto 02 melalui FLP Kids [Motor and cognitive development of students at SD Al-Irsyad Al-Islamiyyah Purwokerto 02 through FLP Kids]. *Jurnal Pendidikan Tambusai*, *8*(2), 19467–19478.

Kusuma, A. S., & Nurmawanti, I. (2023). Pengembangan soal-soal literasi dan numerasi berbasis High Order Thinking Skills (HOTS) untuk siswa sekolah dasar (SD) [Developing literacy and numeracy items based on higher order thinking skills (HOTS) for primary school students]. *Jurnal Ilmiah Profesi Pendidikan*, *8*(1), 516–523.

Kyaruzi, F., & Kisamo, K. (2025). Primary school teachers' competences in constructing quality multiple-choice questions in Tanzania. *International Journal of Evaluation and Research in Education*, *14*(3), 2051–2060. https://doi.org/10.11591/ijere.v14i3.30825

Lestari, W. M., Daryanto, J., & Hadiyah, H. (2023). Analisis kesulitan peserta didik dalam menyelesaikan soal numerasi pecahan pada Asesmen Kompetensi Minimum di sekolah dasar [Analysis of students' difficulties in solving fraction numeracy items in the Minimum Competency Assessment at primary schools]. *Didaktika Dwija Indria*, *11*(1). https://doi.org/10.20961/ddi.v11i1.73462

Marlena, L., Wahidin, W., & Al Azizah, U. S. (2022). Pelatihan kompetensi literasi dan numerasi guru sebagai penguatan menghadapi Kurikulum Merdeka [Training on

teachers' literacy and numeracy competence as reinforcement for facing the Merdeka Curriculum]. *Jumat Pendidikan: Jurnal Pengabdian Masyarakat*, *3*(3), 151–155.

Naeem, M., Ozuem, W., Howell, K., & Ranfagni, S. (2023). A step-by-step process of thematic analysis to develop a conceptual model in qualitative research. *International Journal of Qualitative Methods*, *22*, 1–18. https://doi.org/10.1177/16094069231205789

Ngafifah, K., & Hanif, M. (2026). Pemahaman teks keislaman dan praktik kapital simbolik bahasa Arab di MTs Al-Ittihaad Ma'arif NU 1 Purwokerto [Understanding of Islamic texts and the practice of symbolic capital in Arabic at MTs Al-Ittihaad Ma'arif NU 1 Purwokerto]. *Jumlatuna: Jurnal Pendidikan Bahasa Arab*, *2*(1).

Nugent, G., Houston, J., Kunz, G., & Chen, D. (2023). Analysis of instructional coaching: What, why and how. *International Journal of Mentoring and Coaching in Education*, *12*(4), 402–423. https://doi.org/10.1108/IJMCE-08-2022-0066

Nurwahidah, N., Sari, S. N., Nurmawanti, I., Nisa, K., & Kusuma, A. S. (2023). Pelatihan penyusunan soal HOTS bagi guru-guru SDN 14 Cakranegara [Training in writing HOTS items for teachers of SDN 14 Cakranegara]. *SELAPARANG: Jurnal Pengabdian Masyarakat Berkemajuan*, *7*(2), 1295–1299.

OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

Pastore, S. (2023). Teacher assessment literacy: A systematic review. *Frontiers in Education*, *8*, Article 1217167. https://doi.org/10.3389/feduc.2023.1217167

Pulungan, L. H., & Elfrianto. (2024). Kepelatihan kompetensi dan kapasitas literasi dan numerasi bagi guru SMP di lingkungan Dinas Pendidikan Kabupaten Langkat [Training on competence and capacity in literacy and numeracy for junior high school teachers under the Langkat District Education Office]. *Jurnal Pengabdian Masyarakat Larisma*, *1*(1), 27–35. https://doi.org/10.56495/jpml.v1i1.553

Shamdas, G., Buntu, A., Rafiqa, & Dewi, I. I. (2023). Pelatihan dan bimbingan teknis menyusun assesmen berbasis HOTS pada guru-guru SD Inpres 3 Birobuli [Training and technical guidance in writing HOTS-based assessments for teachers of SD Inpres 3 Birobuli]. *Bubungan Tinggi: Jurnal Pengabdian Masyarakat*, *5*(4), 1566–1573. https://doi.org/10.20527/btjpm.v5i4.9666

Sinta, U. A., Roebyanto, G., & Nuraini, N. L. S. (2022). Analisis kesulitan guru dalam menyusun soal evaluasi berbasis HOTS pada pembelajaran matematika di SDN Torongrejo 2 [Analysis of teachers' difficulties in writing HOTS-based evaluation items in mathematics learning at SDN Torongrejo 2]. *Jurnal Pembelajaran, Bimbingan, dan Pengelolaan Pendidikan*, *2*(1), 45–53.

Ujiyanti, L. N., & Hanif, M. (2025). Evaluasi aspek afektif, kognitif, psikomotorik dalam pembelajaran Pendidikan Agama Islam: Di SMP Negeri 3 Kedungbanteng [Evaluation of affective, cognitive, and psychomotor aspects in Islamic religious education learning: At SMP Negeri 3 Kedungbanteng]. *IQRO: Journal of Islamic Education*, *8*(1), 319–331.