International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/



(RESEARCH ARTICLE)

Check for updates

QUALITY OF MATHEMATICS ASSESSMENT ITEMS BASED ON THE REVISED BLOOM'S TAXONOMY IN UPPER SECONDARY SCHOOLS OF KAMONYI DISTRICT, RWANDA

Christine Nikobahoze * and Egide Masengesho

College of Education, University of Rwanda, Rwanda.
* Corresponding Author

ORCID Details
Christine Nikobahoze: https://orcid.org/0009-0009-7585-7272
Egide Masengesho: https://orcid.org/0009-0005-1393-5563

International Journal of Science and Research Archive, 2026, 21(02), 046–054

Publication history: Received on 21 September 2026; revised on 27 October 2026; accepted on 29 October 2026

Article DOI: https://doi.org/10.30574/ijsra.2026.21.2.1932

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

The cognitive demands of mathematics assessment items influence how effectively assessments represent foundational knowledge, procedural competence, and higher-order thinking skills. This study analyzed the cognitive levels represented in teacher-set mathematics assessment papers from selected upper secondary schools in Kamonyi District, Rwanda. A qualitative content-analysis design was used to examine six archived assessment papers purposively selected from three schools across the 2022–2023 and 2023–2024 academic years. Items were classified into remembering, understanding, applying, analyzing, evaluating, and creating using the Revised Bloom's Taxonomy and a Table of Specifications (TOS). The observed distributions were compared with the benchmark adopted in the study's TOS. Across 74 classified items, applying was the most frequent category, accounting for 52 items (70.3%). Remembering accounted for 9 items (12.2%), understanding for 6 (8.1%), analyzing for 2 (2.7%), evaluating for 3 (4%), and creating for 2 (2.7%). Four papers contained no items classified as analyzing, evaluating, or creating; one contained a single evaluating item; and one included items in all six categories, although their distribution was uneven. The findings indicate that the sampled assessment papers did not consistently represent the full range of cognitive processes specified in the study's TOS. The study recommends strengthening assessment literacy, promoting systematic use of tables of specifications, and providing professional development on constructing mathematics items across a broader range of cognitive demands. Because the analysis covered six papers from three purposively selected schools in one district, the findings describe the selected materials and should not be generalized statistically to all upper secondary schools in Rwanda.

Keywords: Mathematics Assessment, Revised Bloom's Taxonomy, Cognitive Demand, Higher-Order Thinking Skills, Table of Specifications, Rwanda

1 INTRODUCTION

Mathematics education plays an important role in developing learners' logical reasoning, problem-solving abilities, critical thinking, and capacity to apply knowledge to unfamiliar situations. These competencies are relevant to continued education and participation in science, technology, engineering, and mathematics (STEM) fields. Consequently, mathematics assessment should provide evidence of more than students' ability to recall facts and execute familiar procedures. It should also provide opportunities for learners to analyze information, evaluate mathematical arguments, and develop solutions to complex problems (Fitzmaurice et al., 2021; Sachdeva & Eggen, 2021).

Assessment is an essential component of teaching and learning because it helps teachers determine what students understand, identify learning difficulties, and make instructional decisions. However, the educational value of an assessment depends partly on the quality and cognitive demands of its items. An assessment dominated by routine exercises may provide useful evidence of procedural fluency but may offer limited evidence of students' ability to reason, justify conclusions, or solve unfamiliar problems. Effective assessment therefore requires attention to curriculum alignment, learning objectives, content coverage, and the cognitive processes expected of learners (Baranovskaya & Shaforostova, 2017; Sekyi, 2016).

The Revised Bloom's Taxonomy provides a framework for classifying learning objectives and assessment tasks according to six cognitive processes: remembering, understanding, applying, analyzing, evaluating, and creating. Remembering involves retrieving relevant knowledge, while understanding involves explaining or interpreting information. Applying concerns using knowledge or procedures in particular situations. Analyzing involves examining relationships and structures; evaluating requires making judgments based on criteria; and creating involves combining elements to produce a new or coherent solution.

In mathematics, remembering, understanding, and applying are commonly grouped as lower-order thinking skills (LOTS), whereas analyzing, evaluating, and creating are commonly grouped as higher-order thinking skills (HOTS). This distinction can help teachers examine the balance of tasks in an assessment. However, the actual cognitive demand of an item depends on its mathematical content, context, and the reasoning required, rather than on command words alone. Research examining educational materials has used Bloom's Taxonomy to investigate the representation of cognitive processes in learning resources (Qasrawi & Beniabdelrahman, 2020).

Curriculum implementation also depends on the alignment of learning objectives, teaching activities, and assessment practices. A Table of Specifications (TOS) can assist teachers in planning assessments by connecting curriculum content with intended cognitive levels. It can help identify whether the number and types of items represent the content and learning objectives that an assessment is intended to measure. Assessment design should provide opportunities for students to demonstrate meaningful use of knowledge and reasoning rather than relying exclusively on routine performance (Harris et al., 2019).

In Rwanda, the competence-based curriculum emphasizes the development and application of knowledge, skills, attitudes, and competencies. Mathematics assessment should therefore be examined in relation to curriculum objectives and the intended cognitive demands of learning tasks. Nevertheless, curriculum intentions alone do not establish the cognitive demands represented in individual teacher-set assessment papers. Analysis of the actual items is necessary to describe their cognitive distribution.

This study focused on the cognitive categories represented in teacher-set mathematics assessment papers from selected upper secondary schools in Kamonyi District, Rwanda. It examined the items in relation to the Revised Bloom's Taxonomy and the benchmark adopted in the study's TOS.

The purpose of the study was to analyze the cognitive levels represented in six teacher-set mathematics assessment papers from three selected upper secondary schools across the 2022–2023 and 2023–2024 academic years.

The study addressed three research questions:

- How were the items in the selected mathematics assessment papers distributed across the six cognitive categories of the Revised Bloom's Taxonomy?
- To what extent did the observed distributions correspond to the cognitive-level benchmark adopted in the study's TOS?
- What differences in cognitive-level representation were observed among the six assessment papers?

2 LITERATURE REVIEW

2.1 Mathematics assessment and cognitive demand

Mathematics assessment can be used to check knowledge acquisition, monitor progress, identify learning needs, and evaluate achievement of learning objectives. Different assessment items require different forms of thinking. Some require students to recall definitions or reproduce familiar procedures, whereas others require them to interpret information, connect concepts, justify mathematical claims, or develop a solution strategy.

Assessment practices that emphasize routine exercises can provide information about procedural knowledge. However, a narrow range of task types may not adequately represent all the competencies intended by a mathematics curriculum. Assessment design should therefore reflect learning objectives and provide appropriate opportunities for students to demonstrate different forms of mathematical knowledge and reasoning (Sachdeva & Eggen, 2021; Sekyi, 2016).

The relationship between assessment and achievement is complex. Assessment results may be influenced by prior knowledge, instructional practices, learning opportunities, item characteristics, and other contextual factors. Consequently, the cognitive distribution of assessment items should not be treated as a direct measure of student ability or as evidence that a particular assessment design causes a particular level of achievement.

2.2 The Revised Bloom's Taxonomy

The Revised Bloom's Taxonomy classifies cognitive processes into six categories: remembering, understanding, applying, analyzing, evaluating, and creating. It provides a framework for examining the intended thinking processes associated with learning objectives and assessment tasks.

In mathematics, remembering may involve recalling a formula or definition. Understanding may involve explaining a mathematical relationship. Applying may involve using a known procedure to solve a problem. Analyzing may require identifying relationships between expressions or examining a problem's structure. Evaluating may involve assessing the validity of a solution or comparing methods using appropriate criteria. Creating may involve constructing a mathematical model or developing a solution strategy.

These examples are illustrative rather than definitive. Classification requires consideration of the complete item and the reasoning expected from students. An item containing a particular command word does not automatically belong to a specific cognitive category. Analyses of educational materials have used Bloom's Taxonomy to examine the representation of different cognitive levels (Qasrawi & Beniabdelrahman, 2020).

2.3 Tables of Specifications and curriculum alignment

A Table of Specifications is an assessment-planning tool that connects content areas, learning objectives, and cognitive levels. It can help teachers determine whether an assessment represents the intended curriculum and whether the distribution of items reflects the relative importance of different objectives.

A TOS may specify the number or proportion of items allocated to content areas and cognitive categories. Comparing an actual assessment with a TOS can reveal differences between the planned and observed distribution. Interpretation depends on the suitability of the benchmark and the purpose of the assessment. The appropriate cognitive distribution depends on curriculum objectives, learners' level, the content being assessed, and the intended purpose of the assessment.

2.4 Assessment practices and mathematics learning

Assessment practices can influence how teachers organize instruction and how students prepare for examinations. Formative assessment, for example, can provide information that teachers use to identify misconceptions and adapt instruction. Research has examined the relationship between formative assessment, mathematics achievement, and students' attitudes (Kültür & Kutlu, 2021). Other work has investigated higher-order thinking skills, school-based assessment, and mathematics achievement (Azid et al., 2022).

However, findings from different settings cannot automatically be transferred to the schools examined in the present study. Local analysis is needed to describe the characteristics of assessment materials in a particular context. The present study contributes a focused analysis of six teacher-set mathematics assessment papers from three selected upper secondary schools in Kamonyi District. Its purpose is to describe the distribution of cognitive categories in the sampled papers, rather than to measure student achievement or establish causal relationships between assessment practices and learning outcomes.

3 METHODOLOGY

3.1 Research design

The study employed a qualitative content-analysis design to examine the cognitive demands represented in archived teacher-set mathematics assessment papers. Content analysis was appropriate because the study focused on the content and classification of assessment items rather than on participants' experiences, classroom observations, or student responses. Numerical counts and percentages were used to summarize the distribution of coded items and support comparisons among the six papers.

3.2 Study setting and selection of assessment papers

The study was conducted in selected upper secondary schools in Kamonyi District, Rwanda. Three schools were purposively selected from the 30 district schools identified as offering mathematics as a core subject. Six archived teacher-set mathematics assessment papers were selected: one paper from each school for each of the 2022–2023 and 2023–2024 academic years.

The schools were designated School A, School B, and School C to preserve their anonymity in the presentation of findings. Each school contributed two papers. The selection was purposive rather than random and was intended to support an examination of the selected assessment materials. It was not designed to provide a statistically representative estimate for all upper secondary schools in Kamonyi District or Rwanda.

3.3 Data sources

The data consisted of six archived teacher-set mathematics assessment papers. The study examined the questions and sub-questions contained in these documents. No students or teachers were recruited as research participants for this content analysis. The study did not analyze student responses, conduct interviews, or observe classroom teaching. Therefore, the findings concern the cognitive demands represented in the assessment papers and do not directly establish students' mathematical abilities, teachers' intentions, or the effects of assessment design on achievement.

3.4 Analytical framework

The Revised Bloom's Taxonomy was used to classify assessment items into six categories:

- **Remembering:** retrieving or recognizing mathematical facts, definitions, formulas, or procedures.
- **Understanding:** interpreting, explaining, or demonstrating understanding of mathematical concepts and relationships.
- **Applying:** using mathematical knowledge, rules, or procedures to solve problems.
- **Analyzing:** examining mathematical structures, identifying relationships, or separating a problem into meaningful components.
- **Evaluating:** judging the validity of a mathematical argument, method, or solution using appropriate criteria.
- **Creating:** constructing a mathematical model, developing a solution strategy, or combining mathematical ideas to produce a coherent new product.

Classification was based on the cognitive process required by the item as presented, rather than solely on the presence of particular command words.

The study also used a Table of Specifications to compare the observed distribution with the benchmark adopted for the analysis. The benchmark specified 30% remembering, 20% understanding, 20% applying, and 10% each for analyzing, evaluating, and creating. This distribution is treated as the benchmark adopted in the study's TOS, not as an officially mandated national standard.

3.5 Data analysis

Each assessment paper was examined, and its items were classified according to the six cognitive categories. The number of items assigned to each category was recorded separately for each paper. For each paper, the percentage in a category was calculated by dividing the number of items in that category by the total number of classified items in the paper and multiplying by 100.

For the pooled distribution, the number of items in each category across all six papers was divided by the total number of classified items across the six papers. The resulting frequencies and percentages were compared descriptively with the benchmark specified in the TOS. No inferential statistical testing was undertaken.

3.6 Trustworthiness and coding transparency

To ensure the reliability and trustworthiness of the classification process, the Revised Bloom's Taxonomy was applied through a systematic coding protocol. The items across all six assessment papers were independently coded by both authors using the operational definitions of the six cognitive categories. Following independent coding, the authors compared their classifications. Any discrepancies or ambiguous items, particularly those involving overlapping procedural application and reasoning were discussed collaboratively until complete consensus was reached. No formal inter-rater reliability coefficient (such as Cohen's Kappa) was computed, as consensus was achieved for all classified items through iterative review.

3.7 Ethical considerations

The analysis used archived assessment documents and did not recruit human participants or collect student responses. The use of school documents should nevertheless comply with applicable institutional requirements and any conditions imposed by the schools or relevant education authorities.

4 RESULTS

4.1 Distribution of cognitive categories in the six assessment papers

A total of 74 items were classified across the six assessment papers. Table 1 presents the number and percentage of items assigned to each cognitive category in each paper.

Table 1: Number of Items (n) and Within-Paper Percentages (%)by Cognitive Category

School and academic year	Remembering n (%)	Understanding n (%)	Applying n (%)	Analyzing n (%)	Evaluating n (%)	Creating n (%)	Total
School A, 2022–2023	3 (23.1)	2 (15.4)	8 (61.5)	0 (0.0)	0 (0.0)	0 (0.0)	13
School A, 2023–2024	2 (16.7)	0 (0.0)	10 (83.3)	0 (0.0)	0 (0.0)	0 (0.0)	12
School B, 2022–2023	2 (20.0)	2 (20.0)	6 (60.0)	0 (0.0)	0 (0.0)	0 (0.0)	10
School B, 2023–2024	1 (10.0)	0 (0.0)	9 (90.0)	0 (0.0)	0 (0.0)	0 (0.0)	10
School C, 2022–2023	1 (7.1)	2 (14.3)	5 (35.7)	2 (14.3)	2 (14.3)	2 (14.3)	14
School C, 2023–2024	0 (0.0)	0 (0.0)	14 (93.3)	0 (0.0)	1 (6.7)	0 (0.0)	15
Pooled total	9 (12.2)	6 (8.1)	52 (70.3)	2 (2.7)	3 (4)	2 (2.7)	74

Note. Percentages for individual papers were calculated using the total number of items in each paper. Pooled percentages were calculated using the combined total of 74 items. Percentages may not sum to exactly 100 because of rounding.

Applying was the most frequent cognitive category in all six assessment papers. Its representation ranged from 35.7% in School C's 2022–2023 paper to 93.3% in School C's 2023–2024 paper.

Across all six papers, applying accounted for 52 of the 74 items (70.3%). Remembering accounted for 9 items (12.2%), and understanding accounted for 6 items (8.1%). Analyzing, evaluating, and creating together accounted for 7 items (9.5%).

The distribution varied among papers. School A's two papers contained items classified as remembering, understanding, and applying, but none classified as analyzing, evaluating, or creating. Both School B papers showed the same broad pattern. School C's 2022–2023 paper included items in all six categories, while its 2023–2024 paper was dominated by applying and contained one evaluating item.

4.2 Representation of higher-order cognitive categories

Four papers both papers from School A and both papers from School B contained no items classified as analyzing, evaluating, or creating. School C's 2022–2023 paper contained two analyzing items, two evaluating items, and two creating items. Together, these six items represented 42.9% of that paper's 14 items. School C's 2023–2024 paper contained one evaluating item, representing 6.7% of its 15 items, and no items classified as analyzing or creating.

These results show variation in the cognitive categories represented in the selected papers. However, the presence of items in a higher-order category does not, by itself, establish that an assessment adequately measures that cognitive process. The quality, clarity, content alignment, and actual reasoning demands of the items would require further examination.

4.3 Comparison with the Table of Specifications benchmark

The pooled distribution was compared with the benchmark adopted in the study's TOS. Table 2 summarizes the comparison.

Table 2: Comparison Between the Pooled Item Distribution and the Study's TOS Benchmark

Cognitive category	TOS benchmark (%)	Observed pooled percentage (%)	Difference (percentage points)
Remembering	30.0	12.2	−17.8
Understanding	20.0	8.1	−11.9
Applying	20.0	70.3	+50.3
Analyzing	10.0	2.7	−7.3
Evaluating	10.0	4	−6
Creating	10.0	2.7	−7.3
Total	100.0	100.0	0.0

Note. Differences are calculated as the observed pooled percentage minus the benchmark percentage. The benchmark is the distribution adopted in the study's TOS and should not be interpreted as a nationally prescribed standard without supporting documentation.

The pooled distribution differed from the benchmark in all six categories. Applying accounted for 70.3% of classified items, compared with the benchmark of 20.0%. The observed proportions for remembering and understanding were below their respective benchmarks. Analyzing, evaluating, and creating also had lower pooled proportions than specified in the TOS.

These comparisons indicate that the combined set of sampled papers did not match the cognitive distribution adopted for the analysis. However, the pooled results combine six papers from three purposively selected schools. They do not show that every paper differed from the benchmark in the same way or establish the distribution of cognitive demands in all mathematics assessments in the district.

5 DISCUSSION

5.1 Dominance of applying in the sampled assessment papers

The principal finding was the dominance of applying across the six selected mathematics assessment papers. Applying represented 70.3% of the 74 classified items and was the most frequent category in each paper. This distribution suggests that the sampled papers placed considerable emphasis on tasks classified as requiring the use of mathematical knowledge or procedures.

Application is an essential component of mathematics learning. Students need opportunities to use mathematical concepts and procedures accurately in solving problems. Therefore, a high proportion of applying items is not inherently inappropriate. Its significance depends on the intended learning outcomes, the content being assessed, and the balance of cognitive demands required by the relevant curriculum.

Nevertheless, the pooled distribution was substantially higher in applying than the 20% benchmark adopted in the study's TOS. This difference indicates a mismatch between the observed distribution and the benchmark used for

comparison. It provides a reason for reviewing how assessment items are planned and whether the intended balance of cognitive processes is represented in the selected papers.

This interpretation is consistent with the broader view that assessment design should be aligned with learning objectives and should provide evidence of the intended range of knowledge and skills (Baranovskaya & Shaforostova, 2017; Harris et al., 2019). The findings do not establish why applying was dominant. Possible explanations, such as teachers' assessment preferences, time constraints, or familiarity with particular item formats, would require additional data.

5.2 Limited representation of analyzing, evaluating, and creating

Four of the six papers contained no items classified as analyzing, evaluating, or creating. One paper included a single evaluating item, while one included items in all six categories. In the pooled distribution, analyzing, evaluating, and creating accounted for 2.7%, 4%, and 2.7% of items, respectively.

This pattern indicates limited representation of items classified as requiring these higher-order cognitive processes in the sampled materials. In mathematics, such tasks may involve examining relationships, assessing the validity of a solution, comparing strategies, constructing mathematical models, or developing a reasoned approach to an unfamiliar problem.

The finding should be interpreted carefully. Classification under the Revised Bloom's Taxonomy does not automatically determine the quality or difficulty of a mathematical item. Some tasks may require substantial reasoning even when they use familiar procedures, while an item labelled as an evaluation or creation task may not provide strong evidence of the intended process. The present analysis describes the categories assigned to the items; it does not establish students' actual abilities or the effects of the papers on learning outcomes.

The findings nevertheless suggest that the selected papers could be reviewed to determine whether they adequately reflect the range of cognitive processes intended by the study's TOS. Previous research examining higher-order thinking skills and assessment has also investigated the cognitive demands represented in educational tasks (Azid et al., 2022; Qasrawi & Beniabdelrahman, 2020).

5.3 Differences among the selected papers

The papers showed variation in cognitive-category representation. The four papers from Schools A and B contained no items classified as analyzing, evaluating, or creating. School C's 2022–2023 paper included items in all six categories, whereas its 2023–2024 paper was dominated by applying.

This variation suggests that the cognitive distribution was not uniform across the sampled assessment materials. However, the study was not designed to identify the causes of differences among schools or academic years. The small, purposively selected sample does not support broad conclusions about school-level assessment practices or trends over time.

The results are best understood as a description of the six papers examined. Further investigation would be required to determine whether similar patterns occur in a larger sample and whether observed differences are associated with particular assessment-planning practices, curriculum topics, or professional-development experiences.

5.4 Implications for assessment planning and teacher development

The results have practical implications for assessment planning in the sampled schools. Teachers may benefit from systematically linking learning objectives, curriculum content, and item types through a Table of Specifications. Such planning can help identify whether an assessment includes an appropriate range of cognitive demands for its intended purpose.

Professional development could support teachers in distinguishing among cognitive categories and designing mathematics tasks that elicit different kinds of reasoning. Collaborative item review may provide opportunities for teachers to discuss the intended cognitive process of each item, identify ambiguous classifications, and check the alignment of questions with curriculum objectives.

These recommendations should not be interpreted as evidence that the sampled teachers lack assessment competence. The study did not directly examine teachers' knowledge, beliefs, or professional-development experiences. Rather, the recommendations follow from the observed distribution of cognitive categories in the selected assessment papers.

5.5 Limitations and directions for future research

Several limitations should be considered when interpreting the findings. First, the analysis covered only six assessment papers from three purposively selected schools in one district. The results cannot be generalized statistically to all upper secondary schools in Kamonyi District or Rwanda.

Second, the study examined assessment documents rather than student responses or achievement records. It therefore cannot establish whether the cognitive distribution of the papers was associated with student performance or determine how teachers used assessment results to guide instruction.

Third, classification of assessment items requires informed judgment. Although the Revised Bloom's Taxonomy provided a structured framework, some items may involve more than one cognitive process. Clear coding rules and transparent reporting of coder arrangements are important for evaluating the trustworthiness of the analysis.

Fourth, the comparison with the TOS depends on the suitability and provenance of the benchmark. The benchmark should be supported by a clearly identified source or justified as a study-specific analytical framework. Without such justification, the observed differences should be interpreted as comparisons with the benchmark adopted by the researchers rather than evidence of non-compliance with an official standard.

Future research could examine a larger and more diverse sample of assessment papers from additional schools and districts. Studies could also use multiple coders, report appropriate measures of coding agreement where applicable, and examine the relationship between item characteristics and curriculum objectives. Research combining assessment-document analysis with student responses, teacher interviews, or achievement data could investigate questions that the present study was not designed to answer.

6 CONCLUSION

This study analyzed the cognitive demands represented in six teacher-set mathematics assessment papers from three selected upper secondary schools in Kamonyi District, Rwanda, across the 2022–2023 and 2023–2024 academic years. Using the Revised Bloom's Taxonomy, the study classified 74 items into remembering, understanding, applying, analyzing, evaluating, and creating.

Applying was the most frequent category in all six papers and accounted for 70.3% of the pooled items. Four papers contained no items classified as analyzing, evaluating, or creating; one contained a single evaluating item; and one included items from all six categories, although their distribution was uneven. The pooled distribution also differed from the cognitive-level benchmark adopted in the study's Table of Specifications.

The findings indicate that the sampled papers did not consistently represent the full range of cognitive categories specified in the study's TOS. They support reviewing assessment-planning practices and considering professional development focused on assessment literacy, curriculum alignment, and the construction of mathematics items that elicit a broader range of cognitive processes.

The findings are descriptive and apply to the six assessment papers examined. They do not establish students' actual mathematical abilities, explain why the observed distribution occurred, or demonstrate a relationship between assessment design and achievement. Further research using larger and more diverse samples, transparent coding procedures, and additional sources of evidence would help determine whether similar patterns occur in other educational settings.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgments

The authors thank the College of Education, University of Rwanda, for institutional support, and the schools in Kamonyi District that made their archived assessment papers available for this study.

Disclosure of conflict of interest

The authors declare that they have no competing interests that could have influenced the research, its analysis, or the preparation of this manuscript.

Statement of ethical approval

Permission to access the archived mathematics assessment papers was obtained from Kamonyi District and the participating schools. The study analyzed assessment documents and did not recruit students or teachers as research participants.

Author contributions

Christine Nikobahoze conceptualized the study, designed the framework, and drafted the manuscript. Egide Masengesho contributed to the data collection, item classification analysis, and critical revision of the manuscript. Both authors read and approved the final version

Data availability

The coded dataset supporting the findings of this study is available from the corresponding author upon reasonable request, subject to permission requirements and confidentiality considerations. The original mathematics assessment papers are not publicly available because access was granted by Kamonyi District and the participating schools under the applicable permission arrangements.

REFERENCES

[1] Azid, N., Ali, R. M., El Khuluqo, I., Purwanto, S. E., & Susanti, E. N. (2022). Higher order thinking skills, school-based assessment and students' mathematics achievement: Understanding teachers' thoughts. *International Journal of Evaluation and Research in Education, 11*(1), 290–302. https://doi.org/10.11591/ijere.v11i1.22030

[2] Baranovskaya, T., & Shaforostova, V. (2017). Assessment and evaluation techniques. *Journal of Language and Education, 3*(2), 30–38. https://doi.org/10.17323/2411-7390-2017-3-2-30-38

[3] Fitzmaurice, O., O'Meara, N., & Johnson, P. (2021). Highlighting the relevance of mathematics to secondary school students—Why and how. *European Journal of STEM Education, 6*(1), 1–13. https://doi.org/10.20897/ejsteme/10895

[4] Harris, C. J., Krajcik, J. S., Pellegrino, J. W., & DeBarger, A. H. (2019). Designing knowledge-in-use assessments to promote deeper learning. *Educational Measurement: Issues and Practice, 38*(2), 53–67. https://doi.org/10.1111/emip.12253

[5] Kültür, Y. Z., & Kutlu, M. O. (2021). The effect of formative assessment on high school students' mathematics achievement and attitudes. *Journal of Pedagogical Research, 5*(4), 155–171. https://doi.org/10.33902/JPR.2021474302

[6] Qasrawi, R., & Beniabdelrahman, A. (2020). The higher and lower-order thinking skills (HOTS and LOTS) in Unlock English textbooks (1st and 2nd editions) based on Bloom's taxonomy: An analysis study. *International Online Journal of Education and Teaching, 7*(3), 744–758.

[7] Sachdeva, S., & Eggen, P. O. (2021). Learners' critical thinking about learning mathematics. *International Electronic Journal of Mathematics Education, 16*(3), Article em0644. https://doi.org/10.29333/iejme/11003

[8] Sekyi, E. T. A. (2016). Assessment, student learning and classroom practice: A review. *Journal of Education and Practice, 7*(21), 1–6.