Journal of Education, Learning, and Management (JELM)

ISSN: 3079-2541 (Online)

Volume 2 Issue 2, (2025)

https://doi.org/10.69739/jelm.v2i2.1025

https://journals.stecab.com/jelm

Published by
Stecab Publishing

Research Article

Misalignment in Cognitive Levels: A Taxonomic Analysis of Tables of Specification and Mathematics Examinations

*[1]Darryl Jhet P. Bartolome, [1]Jane D. Navalta

About Article

Article History

Submission: July 19, 2025
Acceptance : October 06, 2025
Publication : October 11, 2025

Keywords

Cognitive Level, Table of Specifications, Taxonomic Content Analysis, Test Items

About Author

[1] College of Arts and Sciences, Nueva Vizcaya State University, Bayombong, Nueva Vizcaya, Philippines

Contact @ Darryl Jhet P. Bartolome
djbart178@gmail.com

ABSTRACT

This study examined the alignment between the Table of Specifications (TOS) and periodical examinations in Grade 6 Mathematics at San Mateo West Central School during the 2023–2024 academic year, focusing on the cognitive distribution of test items using Bloom's Revised Taxonomy. Employing a descriptive-comparative design, the research analyzed both the initial teacher-classified TOS and a re-evaluated version through taxonomic content analysis. Findings revealed that teachers initially overestimated the cognitive demand of test items, with a significant emphasis on lower-order thinking skills (LOTS), particularly understanding (56.5%), and minimal representation of higher-order thinking skills (HOTS), such as evaluating (8.5%) and creating (1.5%). Upon reclassification, the majority of test items fell under the applying level (72.5%), with creating questions entirely absent. This discrepancy indicated a misalignment between intended learning outcomes and actual assessments. The study also identified key strengths, such as the emphasis on procedural knowledge, and notable weaknesses, particularly the imbalanced cognitive distribution and lack of HOTS representation. Based on these findings, the study will serve as a basis for assessment program enhancement titled "Strengthening Teachers' Competence in Test Construction" to enhance teachers' ability to classify and construct valid, reliable test items across cognitive levels. The study accentuates the necessity of targeted teacher training and policy reform to promote assessments that better support critical thinking, creativity, and real-world application.

Citation Style:

Bartolome, D. J. P., & Navalta, J. D. (2025). Misalignment in Cognitive Levels: A Taxonomic Analysis of Tables of Specification and Mathematics Examinations. *Journal of Education, Learning, and Management, 2*(2), 194-207. https://doi.org/10.69739/jelm.v2i2.1025

1. INTRODUCTION

Assessment is a crucial element of education, functioning as a means to evaluate student progress, mastery of competencies, and overall learning development (Tarbiyah *et al.*, 2024; Zhou, 2023). This encompasses various methodologies, including traditional examinations, performance- based assessments, and technology- enhanced evaluations (Djamalovna, 2024). Effective assessments practices integrate formative and summative methods, allowing teachers to provide immediate feedback and evaluate long- term student achievement. Despite advancements, aligning assessments with learning objectives remains challenging, as many teachers struggle to develop valid and reliable test items that accurately measure cognitive skills across different levels. To address these misalignments, frameworks such as Bloom's Revised Taxonomy offer a structured approach for evaluating cognitive demand in test construction.

Assessment design faces a critical issue regarding the distribution of cognitive levels. Test items predominantly feature lower order thinking skills such as remembering and understanding while higher order thinking skills, including analyzing, evaluating and creating are inadequately represented (Arta, 2024). This misalignment hinders the development of critical thinking, problem solving, and decision making skills in students, which are essential for success in higher education and practical applications (Thampi, 2023). Moreover comma the preparedness of educators and accessibility of resources continue to pose challenges to the effective execution of violence assessment (Arsyad Arrafii, 2023). Improving assessment literacy and aligning with educational frameworks like Bloom's revised taxonomy allows teachers to create assessment that are more meaningful and equitable, thus enhancing students learning outcomes and overall academic performance. This imbalance is not confined to higher levels of education, but is also evident in basic education, including elementary assessment practices.

Recent studies reveal a persistent disparity in assessment practices, emphasizing a predominant emphasis on lower order thinking skills instead of higher order thinking skills across various subjects and educational levels. An examination of Chinese grade 7 English textbooks indicated that only 26.7% of tasks focus on higher order thinking skills (Honh, 2024). In contrast, Indonesian high school English assessment primarily assesses comprehension with limited attention to advanced cognitive abilities (Utami *et al.*, 2019). Despite continuous efforts to incorporate higher order thinking skills in mathematics education, curricular frameworks the assessments created by teachers continue to primarily focus on lower order thinking skills (Zana *et al.*, 2024). Features encounter significant difficulties in the precise classification and formulation of HOTS based test items, as primary school teachers often misclassify LOTS questions as hats due to a lack of assessment literacy (Driana & Ernawati, 2019). The findings highlight the need for curriculum reforms, targeted teacher training programs, and improved assessment design strategies to foster critical thinking, problem solving, and analytic skills among teacher crystal who are crafting assessments. Given that teachers primarily rely on tools like the Table of Specifications for planning assessments, insufficient training may result in poorly aligned test items across cognitive levels.

Another study shows that many teachers encounter difficulties in correctly classifying test items in alignment with Bloom's revised taxonomy, leading to a predominance of lower order thinking skills in assessment. A common issue observed in the press service mathematics teachers often conflate cognitive levels with problem difficulty, resulting in reliance on alternative classification systems rather than aligning assessments with established taxonomies (Sebastian, 2020). This matter extends beyond mathematics, as textbook examinations revealed ongoing shortcomings in the integration of higher order thinking skills. Recent editions of English textbooks demonstrate a gradual shift towards incorporating higher order thinking skills, despite the continued predominance of comprehension levels questions (Qaswari & Beniabdelrahman, 2020).

Research on teacher created test item in Indonesia's English language assessment reveals that most questions are restricted to the understanding level, with minimal representation of higher order thinking skills beyond analysis (Utami *et al.*, 2019). A comparable pattern is observed in higher education, where final semester examinations in various university courses predominantly emphasize LOTS, with only one course fully adhering to HOTS-based question development (Ginting *et al.*, 2021).

Additionally, teachers frequently encounter challenges in comprehending and applying higher-order thinking skills (HOTS) in assessments, especially in primary education, where test items often emphasize lower-order thinking skills (LOTS) (Driana & Ernawati, 2019). In basic education, while quarterly examinations typically exhibit acceptable reliability, their predominant use of multiple-choice formats constrains the assessment of students' critical thinking and problem-solving skills (Orongan, 2020).

Recent studies indicate a gap in the assessment of cognitive skills within diverse educational settings, highlighting a predominant focus on Lower Order Thinking Skills (LOTS) at the expense of Higher Order Thinking Skills (HOTS). In Chinese Grade 7 English textbooks, merely 26.7% of tasks were identified as higher-order thinking skills, with the predominant emphasis on basic recall and comprehension (Hong, 2024). A similar trend was observed in Indonesian 12th grade history textbook, where specific materials contained up to 96.7% lower order thinking skills based questions (Najuah *et al.*, 2024). An examination of 7th grade Indonesian EFL textbooks indicated that 80.4% of reading comprehension questions assess lower order thinking skills instead of promoting higher order analytical capabilities (Maryamah *et al.*, 2024). Research of Indonesian high school English test revealed a trend in a teacher designed assessment characterized by a predominance of understanding level questions, with limited inclusion of higher order thinking skills based items (Uta *et al.*, 2019).

According to Ginting *et al.* (2021) Noted that most final semester university combination lock assessments focus on higher order thinking skills come out with only one course fully adopting HOTS- based question writing. The results reveal a notable difference in cognitive classification accuracy across different educational levels, highlighting the need for targeted training or frameworks like Blooms revised taxonomy.

This study examines the table of specification (TOS) and the cognitive levels of test items in periodical examinations to evaluate their alignment with blooms revised taxonomy. Discrepancies, its strengths and weakness was also identified on this research that examines the grade 6 mathematics periodical tests in San Mateo West Central School focusing on the distribution of cognitive domains within test construction practices during the school year 2023-2024.

2. LITERATURE REVIEW

2.1. Revised blooms taxonomy

Bloom's taxonomy, developed by Benjamin bloom in 1956, serves as a fundamental framework for categorizing educational objectives based on cognitive complexity levels the taxonomy provides a systematic framework for curriculum design, instruction, and assessment, organizing learning into six hierarchical levels: Knowledge, comprehension, application, analysis, synthesis, and evaluation (Bloom *et al*, 1956). The framework is widely employed across various disciplines to structure learning outcomes and ensure that assessments align with cognitive processes. Evolution of educational practices highlighted the limitations of the original taxonomy, prompting revisions to a better align with modern learning theorists and instructional methodologies (Krathwohl, 2022).

Anderson and Krathwohl (2001) Revised Bloom's taxonomy to meet changing educational needs through the redefinition and restructuring of its cognitive domains. The revision replaces static nouns with active verbs emphasizing the dynamic aspects of learning process as illustrated in figure 2. The updated hierarchy comprises remembering, understanding, applying, analyzing, evaluating, and creating, with the apex level shifting from synthesis to creation (Anderson & Krathwohl, 2001).

This modification underscores the importance of fostering higher order thinking skills in education as recent studies demonstrate the need for students to develop critical thinking, problem solving and creativity (Dwyer *et al.*, 2014). The updated taxonomy provides a pertinent framework for teachers regarding curriculum design, assessment development and the facilitation of meaningful learning experiences that prepare students for complex real-world challenges (Forehand, 2010).

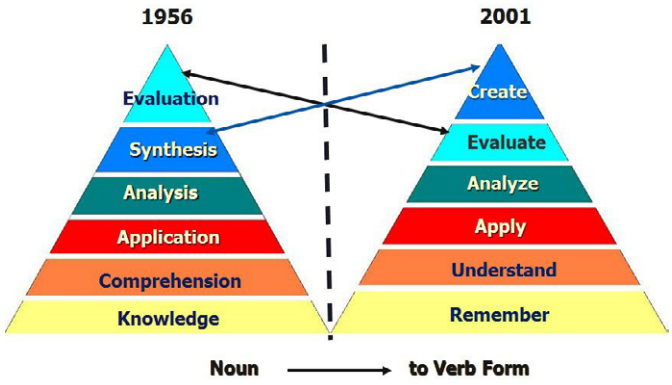

Figure 1. Revised blooms taxonomy

The Revised Bloom's Taxonomy classifies cognitive skills into two main categories: lower order thinking skills (LOTS) and higher order thinking skills (HOTS). This framework ensures student engagement in diverse cognitive processes through a systematic method for developing instructional objectives and assessments. Teachers must understand these levels to align instructional strategies with the intended learning outcomes (Suryani *et al.*, 2020).

LOTS consists of the essential cognitive levels: remembering, understanding, and applying. The process of remembering involves recalling facts, concepts, or procedures without necessarily comprehending them period to comprehend one must articulate concepts, summarized information, or interpret ideas in personal terms. Application extends beyond mere understanding and requires the use of acquired knowledge in new or practical context (Saputra *et al.*, 2022). Although these cognitive levels are crucial for acquiring foundational knowledge, research suggests that an overemphasis on lower order thinking skills in assessment may hinder students' ability to participate in more advanced cognitive processing and problem solving (Setiawan *et al.*, 2019).

Conversely, HOTS include the higher order cognitive processes: analyzing, evaluating, and creating. Decomposing information into components is essential for analyzing relationships and patterns during the analytical process. Evaluating involves creating assessments based on established criteria, such as assessing the effectiveness of a solution or analyzing arguments. The creation of high-level outputs necessitates the synthesis of knowledge, generation of new ideas, and design of innovative solutions (Gunawan *et al.*, 2021). Studies demonstrate that students engage in HOTS based learning activities develop enhance critical thinking, adaptability, and decision-making skills which are essential for education in the 21st century (Putra *et al.*, 2020).

Ensuring a balance between instructional design and assessment is crucial for fostering comprehensive student learning. This encompasses the integration of both lower order thinking skills and higher order thinking skills. While foundational knowledge is essential, prioritizing higher order thinking skills and tests construction enhances critical thinking, problem solving of complex issues and the creation of innovative solutions in practical context. Integrating higher order thinking questions in two periodical tests, teachers can equip students with essential skills necessary for success in higher education and beyond (Widana *et al.*, 2019).

Revised Bloom's Taxonomy serves as a critical framework for that thorough assessment of students learning aligning educational objectives with cognitive process. Teachers often encounter challenges incorrectly categorizing test items according to the revised Bloom's taxonomy. Misconceptions can occur, leading to an overemphasis on lower order thinking skills and insufficient incorporation of higher order thinking skills in periodical tests. This imbalance may hinder the development of higher cognitive abilities in pupils' period to address these challenges, it is crucial to provide teachers with resources and professional development tailored to enhance their assessment literacy. Improving educational outcomes requires enhancing teachers understanding of cognitive classifications, allowing for the development of assessments that more effectively promote and measure higher order thinking skills.

Ensuring alignment between test items and intended learning outcome is crucial for accurately assessing students' cognitive abilities and mastery of competencies. When test items do not align with learning objectives, students may face assessments that inadequately challenge them or fail to accurately evaluate their understanding may leading to disparities in skilled development. Research indicates that assessment focusing predominantly on lower order thinking skills does not effectively promote critical thinking, deep learning or problem-solving abilities (Retnawati *et al*., 2018). The integration higher order thinking skills in periodical test enhances the learning experience by prompting students to engage in analysis, evaluation, and the creation period to achieve this alignment teacher should utilize cognitive frameworks, such as Bloom's revised taxonomy for the development of test items that are valid and reliable (Pratama & Retnawati, 2018).

The common misclassification of test items and excessive dependence on recall-based questions stem from the insufficient formal training of many teachers in developing assessment based on higher order thinking skills (Retnawati *et al*., 2018). Structured capacity building programs, workshops and collaborative assessment design initiatives are essential to address this issue. Furthermore, to ensure thorough evaluations educational institutions and agencies must adopt policy intervention that mandates assessment frameworks incorporating both lower order thinking skills and higher order thinking skills. The development of teachers' competences in test construction enhances instructional practices, improves learning outcomes and prepares students for complex cognitive challenges in their academic and professional pursuits.

2.2. Principles of test construction

An effective assessment is essential for teaching and learning, functioning to evaluate students' mastery of learning objectives, offer feedback for instructional enhancement, and impact educational policy decisions (Brookhart, 2018). Validity is attained when assessments accurately reflect the intended learning outcomes. Reliability is established through consistent results across various administrations. Fairness is ensured by minimizing bias and providing all students with equal opportunities to showcase their knowledge. Practically is achieved by balancing feasibility concerning time, resources, and implementation (Bandalos, 2018).

Research demonstrates that incorporating these principles into assessment design led to more significant evaluation of students performance, thus enhancing their function and facilitating learning and aligning with the curriculum (Newton & Shaw, 2018). Nonetheless, the importance of professional development programs aimed at improving teachers abilities to create the reliable and valid assessment is highlighted by the considerable challenges many face in the test construction process due to the insufficient training in assessment literacy (Gotch & French, 2020). Ensuring that assessment adheres to fundamental principles enhances the overall quality of education, thereby strengthening instructional practices.

Accurate measurement of learning competencies by test items is crucial for conducting effective assessments. Reliability refers to the consistency and stability of test results across multiple administrations, whereas validity denotes the extent to which a test accurately measures it intends to measure (Newton & Shaw, 2018). An effective assessment should align with instructional objectives and curriculum standards, ensuring that test items accurately reflect the intended learning outcomes (Bandalos, 2018). There are several threats to the validity and reliability of test items, such as ambiguous wording, test bias, and poorly constructed items, which can distort student performance and misinterpret their actual abilities (AERA *et al*., 2018). Moreover reliability may be adversely affected by inconsistencies in scoring procedures and environmental factors present during test administration (Brookhart, 2018).

To enhance the reliability and validity of assessment, it is advisable to align test items with learning objectives, conduct pilot test before full implementation, and employ multiple assessment methods to triangulate students' performance (Gotch & French, 2020). Furthermore, teachers may improve assessment quality throughout the incorporation of statistical analysis, adherence to standardized tests administration protocols and the use of rubrics (Johnson ert al., 2021). To equip teachers with the necessary skills for creating assessments that are valid and reliable, participation in teacher training and professional development program focused on assessment literacy is crucial (Xu & Brown, 2018). Implementing these principles schools can ensure that their assessment accurately measures students learning and yield meaningful data to inform instruction.

The table of specification is crucial for aligning test items with learning competencies, thus providing a systematic framework for assessment development. Teachers can achieve a balance between lower order thinking skills and higher order thinking skills by strategically distributing test items across different cognitive levels through a well-constructed (TOS) as detailed in Bloom's revised taxonomy (Anderson & Krathwohl, 2001). Research indicates that an over representation of lower order thinking skills often results from a misalignment between the intended outcomes of assessment and actual evaluations, whereas higher order thinking skills are frequently underrepresented (Baird *et al*., 2019). This imbalance limits student's ability to engage in critical thinking and problem solving, which are essential for practical applications.

According to Herman and Linn (2020), many assessments do not accurately reflect the intended learning outcome. Research indicates that classroom based assessments often led to misclassification of questions by teachers, which result in examination that fails to effectively assess the intended competencies (Gareis & Grant, 2020) identified inconsistencies highlight the need for capacity building program aimed at enhancing teachers abilities to create assessment that effectively evaluate students learning. This misclassification of cognitive levels in test items is a significant issue in test construction, as teachers frequently find it challenging to distinguish between lower order thinking skills and higher order thinking skills (Sebastian, 2020).

Research indicates that teachers often categorize remembering and understanding questions as higher order thinking questions, resulting in a misrepresentation of student's cognitive capabilities (Utami *et al*., 2019). This limitation

restricts students' opportunities for deeper learning experience and impedes the development of essential competencies for the 21st century (Ginting *et al*., 2021). Moreover, the processes of hiding writing and question formulation presents further challenges, as teachers frequently lack formal training in the construction of valid, reliable, and equitable test items, resulting in ambiguous or misleading questions.

3. METHODOLOGY

3.1. Research method

This study employed a Multi- method Approach specifically, descriptive- comparative research design to examine the alignment between the Table of Specifications (TOS) and periodical examinations among Grade 6 pupils of San Mateo West Central School for the school year 2023-2024. The Descriptive aspect involved analyzing the distribution of cognitive levels in the test items based on the Revised Blooms Taxonomy, while the comparative components is the content analysis focused on identifying discrepancies between teachers' initial classifications of test items and the result of the taxonomic content analysis.

Quantitative data were gathered by reviewing the TOS and corresponding periodical examinations across four quarters. The cognitive levels of test items were categorized into remembering, Understanding, Applying, Analyzing, Evaluating, and Creating. Statistical tools were used to measure the frequency counts and percentage distribution of these cognitive levels, allowing for an objective comparison through content analysis.

3.2. Sources of data

This study utilized a 200- items periodical test papers and their corresponding Tables of Specification (TOS) for Grade 6 Mathematics across four quarters with 50- items in each quarter in the school year 2023-2024, developed by the District Mathematics Master Teacher Coordinator as the primary sources of data. The District Mathematics School Head Coordinator conducted a comprehensive review of these assessment materials to verify the content validity and alignment to the Most Essential Learning Competencies (MELCs). The final versions received approval from the Principal In- charge. The gathered instrument was used as the basis for the taxonomic content analysis, facilitating a thorough assessment of the distribution of cognitive levels among test items and their correspondents with the TOS.

3.3. Sampling design of the study

This study utilized a purposive sampling design to select the test papers and tables of specifications (TOS) utilized in the periodical examinations of Grade 6 Mathematics at San Mateo West Central School for the school year 2023-2024. Since the researcher is a Mathematics teacher, this subject was deliberately chosen to align with the researcher's field of expertise, ensuring a through and informed analysis of the cognitive levels of test items based on the Revised Bloom's Taxonomy.

3.4. Data gathering procedure

Initially, formal permission was sought from the school principal of San Mateo West Central School to access the Grade 6 Mathematics periodical examination and their table of specifications for the school year 2023-2024. Upon securing approval, copies of the test papers and TOS was obtained from the school testing coordinator.

The test papers and TOS were systematically analyzed and categorized based on the taxonomic levels specified in the Revised Bloom's Taxonomy: Remembering, Understanding, Applying, Analyzing, Evaluating and Creating. The initial phase of the analysis entailed documenting the teacher's classification of test items according to their TOS. A re-evaluation was conducted using taxonomic content analysis to determine the cognitive levels of the test items. The objective of this process was to identify discrepancies between the initial classification of teachers on test items on TOS and those determined through objective assessment of cognitive levels.

The results from the initial and re- evaluated classifications were documented, establishing a foundation for identifying the strengths and weaknesses in the test construction after the content analysis.

The collected data informed the development of a proposed enhancement program for assessment, with the objective of improving the alignment of test items and table of specifications to the Most Essential Learning Competencies (MELCs).

3.5. Data analysis

This study employed both quantitative and qualitative data analysis methods to examine the alignment between the periodical test items and TOS in Mathematics for Grade 6 pupils at San Mateo West Central School.

The following are the statistical tool used on the study:

3.5.1. Descriptive statistics

• *Frequency Counts and Percentages:* These were used to determine the distribution of cognitive level of test items based on the Revised Bloom's Taxonomy, as initially classified by teachers and re- evaluated through taxonomic content analysis,

3.5.2. Qualitative analysis

Themes: The study identified recurring patterns or themes regarding the discrepancies in the initial TOS and the strengths and weaknesses of the constructed TOS.

• *Taxonomic Content Analysis:* This was utilized to re- examine and classify the cognitive levels of the test items, allowing for an objective comparison between teacher's initial classification and the re- evaluated TOS.

As the researcher concurrently serves as a Mathematics teacher within the same district where the study was conducted, it is imperative to acknowledge the potential for researcher bias that may arise from this insider position. Recognizing this, deliberate steps were undertaken to minimize subjectivity and ensure methodological rigor in both the classification and interpretation of data. The researcher adhered strictly to the cognitive frameworks established under Bloom's Revised Taxonomy to guide the content analysis of test items and Tables of Specification (TOS). To enhance credibility, inter-rater validation was conducted by two independent experts in educational assessment who reviewed and confirmed the

accuracy of the cognitive classifications. These measures collectively safeguarded the objectivity of the study, ensuring that findings were grounded in empirical evidence rather than personal familiarity with the research context.

4. RESULTS AND DISCUSSION

Table 1. Cognitive levels of test items distributed in the Table of Specifications (TOS) and periodical examinations as initially classified by teachers

Periodical Test	Taxonomic Level					
	Remembering	Understanding	Applying	Analyzing	Evaluating	Creating
Quarter 1	1	29	8	7	2	3
%	2.00	58.00	16.00	14.00	4.00	6.00
Quarter 2	0	30	9	6	5	0
%	0.00	60.00	18.00	12.00	10.00	0.00
Quarter 3	6	24	7	8	5	0
%	12.00	48.00	14.00	16.00	10.00	0.00
Quarter 4	0	30	7	8	5	0
%	0.00	60.00	14.00	16.00	10.00	0.00
Grand Total	7	113	31	29	17	3
%	3.50	56.50	15.50	14.50	8.50	1.50

Table 1 presents the cognitive levels of test items distributed in the Table of Specifications (TOS) and periodical examinations, as initially classified by teachers, reveal a distinct imbalance across the six levels of the Revised Bloom's Taxonomy. The data demonstrates a strong emphasis toward lower-order thinking skills (LOTS), particularly in the remembering, understanding, and applying categories, while higher-order thinking skills (HOTS)—analyzing, evaluating, and creating—remain noticeably underrepresented. This imbalance raises concerns regarding the depth and complexity of assessments in measuring students' cognitive abilities.

When considering the cumulative data across all quarters, understanding-level questions overwhelmingly accounted for 113 items or 56.50% of the total test items, followed by 31 applying questions at 15.50% and analyzing at 14.50% (29 items). There were 17 evaluating questions made up 8.50%, while remembering constituted just 3.50% or 7 items, and creating accounted the least items with 3 questions or 1.50%. These figures reveal a clear overemphasis on LOTS, with understanding dominating the cognitive distribution, and an alarming underrepresentation of HOTS—particularly creating, which was nearly non-existent.

In the first quarter, the majority of test items focused on the understanding level, comprising 58% of the total, followed by applying at 16% and analyzing at 14%. Evaluating and creating were represented by only 4% and 6%, respectively while remembering accounted for a mere 2%. The heavy concentration of items at the understanding level indicates a predominant reliance on comprehension-based questions, offering limited opportunities for students to engage in critical thinking or creative problem-solving.

The trend persisted in the second quarter, with understanding-level items increasing to 60%, further solidifying the focus on LOTS. Applying and analyzing levels were represented by 18% and 12% of the items, respectively. Although evaluating questions rose slightly to 10%, remembering and creating was entirely excluded (0%). The lack of remembering and creating level items suggests a missed opportunity to cultivate innovation, synthesis, and original thought among learners.

In the third quarter, understanding continued to dominate at 48%, though there was a slight increase in remembering, rising to 12%. Applying (14%) and analyzing (16%) remained moderately represented, while evaluating questions stayed at 10%, and creating was once again absent. While the small rise in analyzing questions hints at an effort to integrate HOTS, the overall focus remained heavily skewed towards LOTS.

The fourth quarter mirrored the patterns observed in the second quarter, with understanding items maintaining a strong presence at 60%. Applying and analyzing questions accounted for 14% and 16%, respectively, while evaluating comprised 10%. Notably, remembering and creating questions were, yet again, excluded.

The findings in Table 1 demonstrate a notable reliance on lower order thinking skills, particularly at the understanding level, across all quarters, with a minimal occurrence of higher order thinking skills including analyzing, evaluating, and creating questions period the observed imbalance indicates that the test items are originally categorized by teachers, may not adequately prompt students to participate in complex cognitive processes (Chew & Cerbin, 2021).

These patterns suggest a misalignment between the construction of test items and the objectives of prompting critical thinking, problem solving and creativity skills vital for deeper learning (Eswaran, 2024). The prevalence of understanding level questions indicate that teachers may prioritize assessing students abilities to recall and comprehend information's rather than their capacity to apply, analyze, or create base on their learning. This may lead to instructional practices

that prioritize route learning and superficial comprehension, rather than fostering critical and innovative thinking among students (Kotsis, 2024). The restricted application of higher order thinking skills and assessment raises concerns regarding the adequacy of student's preparation for real world problem solving an advanced academic challenges. If assessments do not reflect the full range of cognitive levels outlined in Bloom's Taxonomy, there is a risk that classroom instruction might mirror this gap, reinforcing passive learning rather than active engagement with content (Sohdi, 2025).

Table 2. Cognitive levels of the same test items distributed in the TOS after a taxonomic content analysis

Periodical Test	Taxonomic Level					
	Remembering	Understanding	Applying	Analyzing	Evaluating	Creating
Quarter 1	0	4	46	0	0	0
%	0.00	8.00	92.00	0.00	0.00	0.00
Quarter 2	0	14	33	3	0	0
%	0.00	28.00	66.00	6.00	0.00	0.00
Quarter 3	7	8	32	3	0	0
%	14.00	16.00	64.00	6.00	0.00	0.00
Quarter 4	2	11	34	2	1	0
%	4.00	22.00	68.00	4.00	2.00	0.00
Grand Total	9	37	145	8	1	0
%	4.50	18.50	72.50	4.00	0.50	0.00

The frequency and percentage distribution of the cognitive levels of test items distributed in the Table of Specifications (TOS), following a taxonomic content analysis is shown in table 2. The above table reveal a significant shift in the classification of test items across Revised Bloom's Taxonomy. The reclassification highlights a strong concentration of test items at the applying level, with minimal representation of higher-order thinking skills (HOTS) and a notable absence of creating-level questions across all quarters.

Based on the grand total, applying-level items accounted for an overwhelming 72.50% or 145 questions of all test items, indicating a predominant focus on procedural knowledge and the practical application of concepts. Understanding-level questions followed at 37 items or 18.50%, while remembering items constituted only 9 items or 4.50%. Analyzing-level questions were limited to 8 items or 4.00%, with evaluating at a mere 1 item or 0.50%, and creating questions entirely absent (0.00%). This distribution reveals a marked imbalance, heavily favoring mid-level cognitive skills, particularly applying, while neglecting the more complex cognitive processes of evaluating and creating.

In the first quarter, applying-level questions dominated at 92%, with understanding items making up the remaining 8%. Interestingly, there were no items classified under remembering, analyzing, evaluating, or creating. This extreme focus on application suggests an intent to assess procedural knowledge but lacks the necessary cognitive diversity to engage students in either foundational recall or higher-level critical thinking.

The second quarter presented a slight increase in cognitive variety, although applying-level questions still comprised 66% of the total. Understanding questions accounted for 28%, while analyzing appeared at 6%—the first instance of a HOTS category in the data. However, remembering, evaluating, and creating items were entirely absent, reinforcing the pattern of assessments being concentrated on procedural tasks with limited opportunities for students to engage in critical analysis or evaluative judgment.

In the third quarter, applying-level items remained dominant at 64%, but there was a slight increase in the distribution of lower and higher cognitive levels. Remembering questions appeared at 14%, and understanding at 16%, while analyzing made up 6%. Despite this slight broadening of cognitive levels, evaluating and creating were still not represented, pointing to a continued lack of emphasis on complex cognitive processes.

The fourth quarter showed a modest diversification of cognitive levels, with applying items at 68%, understanding at 22%, remembering at 4%, and analyzing at 4%. Evaluating questions appeared for the first time, although at a minimal 2%, while creating-level questions remained absent. Though this quarter exhibited the most balanced distribution, the figures still reflect an overarching focus on mid-level cognitive tasks, with higher-order thinking receiving minimal attention.

The result of the data indicates a consistent pattern of test items being heavily skewed towards the applying level, with limited representation of other cognitive levels especially HOTS. As mention by Kouicem and Boulekhal (2025), the complete absence of creating-level questions across all quarters suggests a critical gap in fostering innovation and synthesis among students.

This reclassification indicates a misalignment between teachers' perceptions and the actual cognitive levels of their test items. While teachers initially believed their assessments contained a balanced representation across Bloom's Taxonomy, the taxonomic content analysis exposes an overemphasis on applying-level tasks. This suggests that the test items primarily evaluated students' capacity to apply learned concepts in familiar

contexts, rather than fostering deep critical thinking or creative problem-solving, which reflects a mid-level cognitive skill.

This imbalance has implications for instructional assessment. The lack of evaluating and creating questions indicates a missed opportunity to enhance students' capacity to criticize ideas, make judgments, and generate new concepts key elements of higher order thinking (Baddane & Ennam, 2024). As a result, instructional practices may exhibit these trends, prioritizing procedural knowledge and overlooking the cultivation of analytical and creative skills. These may limit students' capacity to tackle complex real-world problems or to participate in independent and innovative thinking.

4.1.Discrepancies exist between the teachers' initial classifications and the re-evaluated cognitive levels of test items on the tos

When assessing students learning outcomes, it is crucial that test items effectively reflect the cognitive requirements of each competency. The table of specifications serve as a critical tool for ensuring that assessment items are aligned with learning objectives and cognitive levels. The following discrepancies were identified in the taxonomic content analysis organized through themes.

4.2. Underestimation of higher-order thinking skills (HOTS)

A notable trend identified through the taxonomic content analysis is the insufficient emphasis on higher order thinking skills in the teacher's original table of specification. This contrasts with Miterianifa *et al.*, (2021), which asserts that higher order thinking questions are vital in contemporary education, aligning with the 21st century learning requirements and fostering enhanced cognitive engagement among students the integration of higher order thinking skills into periodical examination allowed teachers to assess students mastery of content while promoting critical thinking, creativity, and problem solving abilities, essential for academic progress and future successes (Wibowo *et al.*, 2024).

The revised TOS indicates a substantial modification as many test questions formerly categorized under lower order thinking skills specifically remembering and understanding have been reclassified into higher cognitive domains including applying analyzing and evaluating. These misconceptions reflect a tendency among teachers to perceive problem solving tasks and multi-step computation merely as comprehension exercises, rather than recognizing the intricate cognitive skills involved. Research consistently shows that lower order thinking skills, particularly those focused on remembering and understanding, dominate the type of questions used in examinations (Atna *et al.*, 2022).

Despite the increasing focus on misalignment of taxonomic levels to its test items and cultivating critical thinking, many teachers find it challenging to formulate HOTS questions proficiently. A study indicated merely 7.5% of the questions devised by teachers were to the domains of analyzing, evaluating or creating shows a significant gap as to its alignment to test items (Musliha *et al.*, 2021). Moreover, pre- service teachers frequently interpret LOTS and HOTS based on question difficulty rather than the cognitive processes they aim to address, indicating a necessity for enhance pedagogical training in developing assessments that effectively measures HOTS and alignment to the levels of Blooms Revised Taxonomy.

The result of the study has continuously demonstrated that teachers often overrate the cognitive complexity of their assessment items, frequently miscategorizing lower-order questions as higher-order ones (Prihastuti & Widodo, 2019; Prihastuti *et al.*, 2020). This under representations of HOTS questions not only limits students ability to develop critical thinking, problem solving, and creative abilities but only perpetuates a cycle where examinations primarily test lower order thinking skills (LOTS) (Faradella, 2024). As a result, students may become acquainted to rote learning practices, relying on memorization rather than genuine engagement with the subject matter (de Sousa, 2023). The emphasis on LOTS questions weakens students' capacity to apply knowledge to new context as a crucial skill in practical problem solving.

This imbalance impacts on students and obstructs efforts to develop critical thinking and problem-solving skills. An analysis of various teachers generated assessments revealed a significant dominance of questions targeting lower order cognitive skills, including recall and comprehension, while there was conspicuous lack of item focus on higher order skills such as analysis evaluation and synthesis (Prihastuti *et al.*, 2020; Abosalem, 2016).

4.3. Misalignment of cognitive levels

A recurring theme is the clear discrepancy between the desired learning outcomes and the cognitive requirements of the test items. The re- evaluated TOS indicates that numerous competencies, particularly those related to problem-solving with decimals, fractions, and word problems, were misclassified as "Understanding," despite their degree of complexity necessitating "Applying" or "Analyzing" skills. This misalignment indicates an overlap of cognitive processes, where tasks necessitating strategy implementation, inference drawing, and logical reasoning were oversimplified in their cognitive classification.

Research of the alignment of assessments with learning objectives, utilizing Revised Bloom's Taxonomy, indicate a recurring discrepancy between its cognitive level and test items. Trevisan and Amaral (2016) states that assessments, frequently emphasize lower cognitive level, such as remembering and understanding, while overlooking higher order competencies, including analyzing, evaluating and creating. The exam itms in mathematics mostly assessed application skills with minimal emphasis on analyzing and evaluating (Basol *et al.*, 2016).

The misalignment of cognitive levels in test and TOS may arise from teachers insufficient comprehension of how to formulate test items that corresponds with the more complex cognitive domains of Revised Blooms Taxonomy. This is consistent with Choy *et al.* (2009) which mentioned that teachers mistakenly assigned taxonomic level in creating examinations and design tests that prioritize factual memory and fundamental knowledge. Moreover, teachers who are on the services for quiet long frequently find it challenging to distinguish the different levels of Blooms Taxonomy, conflating Higher Order Thinking Skills

(HOTS) with the simple complexity of a question instead of the cognitive process included (Sebastian, 2020). The disparity in cognitive level will misrepresent pupils actual intellectual abilities and obstructs the cultivation of essential problem solving and analytical skills.

Fostering students' higher-order thinking capabilities necessitates a collaborative, interdisciplinary methodology, guaranteeing the gradual enhancement of these cognitive competencies throughout their educational trajectory (Abosalem, 2016). The use of cognitive frameworks such as Bloom's revised taxonomy for item classification provides a structured method for creating balance assessments that effectively cover diverse cognitive processes, thereby improving pertinence learning experiences (Dagostino *et al.*, 2014).

4.4. Strengths and weaknesses of the constructed table of specification (tos) and test items

The following are the strengths and weaknesses of the constructed Table of Specifications (TOS) and Test Items. The discussion is divided in to two parts, the strengths and weakness's part. Each part composed of themes to discuss the salient patterns and trends present on the constructed TOS.

4.5. Strengths

4.5.1. Strong Emphasis on Skill- based Assessment and Practical Application

After the taxonomic content analysis, the newly develop TOS clearly emphasizes on the application level of the Revised Bloom's Taxonomy accounting to 72.50% of the test items where in students are required to apply their acquired knowledge in real- world scenarios. This modification shows a deliberate effort to develop test items that extends mere recall and comprehension, emphasizing procedural knowledge and the ability to answer problem solving. This approach is parallel with the objective of improving students' ability to utilize their acquired knowledge in real- world setting.

Based on the findings of this study, which is also align with the result of the study conducted by Harries *et al.* (2019), which advocates for an organized approach in facilitating assessments in the classrooms, putting emphasis on the importance of assessing pupils' ability to apply acquired knowledge based on the most appropriate context. This technique stresses that aligning test items when assessing students with higher- order cognitive skills, where learners are not just memorizing facts or information's but are demonstrating their ability to synthesize, analyze and evaluate data. This concept is furthered by Nilimaa (2023), which encourages students to formulate and solve their own problem mathematically. With this method in mind, students are able to evaluates their procedural knowledge while fostering independent thinking and creativity which are both important component in the complex challenges posed in the real- world context.

Richter-Beuschel *et al.* (2018) proposed a new approach in evaluating procedural knowledge in consonance with the sustainable development, utilizing a combination of think-aloud studies and Delphi method to assessed students' cognitive processes comprehensively. Their study points out the relevance of assessments that focus on HOTS and how students apply concepts in the real- world situations. Barber *et al.*, (2015), examines the relationships between authentic assessments, problem- based learning and collaborative communities in digital environments. This highlights the importance of empowering students and enabling them to choose how they show off their acquired knowledge. This reinforces those assessments should provide opportunities for students to demonstrates their creative and problem-solving abilities.

The focus on application-level questions ensures that students have both conceptual understanding and the capability to executes tasks, connect ideas and solve problem in a more advance and creative way rather that a simple memorization.

4.6. Structured cognitive mapping for balanced assessment

The updated TOS categorizes test items to an accurate and precise cognitive level, creating a clear and organized framework for assessment design. This systematic mapping improves the clarity of learning outcome assessments and ensures alignment between assessment methods and instructional objectives. The majority of the test items primarily emphasizes the applying level; however, the inclusion of other cognitive domains, such as 18.50% are under understanding and 4% and under analyzing level, reflects a low to moderate knowledge on aligning test items to correct cognitive level.

This aligns with the findings of Utam *et al.*, (2019), which indicate that teachers often encounter difficulties in developing test items that adequately assess HOTS, as most questions tend to focus primarily under LOTS, particularly those questions related to understanding level. Many teachers acknowledge the importance of integrating HOTS into designing test items; however, they encounter difficulties in creating reliable, valid and objective test instruments that align with student's primary need. The disparity between intent and implementation suggests that, while the significance of HOTS is recognized, the practical skills and knowledge are necessary for developing effective assessments remain insufficiently developed among teachers.

Furthermore, a study conducted by Driana and Ernawati (2019) illustrates a significant misunderstanding: teachers frequently misidentify lower-order thinking questions as higher-order thinking skills, indicating a constrained grasp of cognitive levels. In response to these challenges, new assessment tools have been implemented, including the three-tier diagnostic test created by Sari and Sunyono (2019), designed to effectively evaluate students' higher-order thinking skills and minimize conceptual misunderstandings. The findings highlight the critical necessity for comprehensive teacher training and ongoing professional development aimed at the design and execution of higher-order thinking skills assessments. Enhancing teachers' ability to develop well-designed test items will improve the validity of assessments and promote students' critical, analytical, and creative thinking vital skills for addressing complex, real-world challenges.

4.7. Weaknesses

4.7.1. Imbalance in cognitive level distribution

The newly developed TOS demonstrates an imbalanced

focus on the application of cognitive level, with minimal representation of higher-order thinking skills (HOTS) such as analysis, evaluation, and creating. Application-based exam items constituted 72.50% of the total, whereas the levels of analyzing (4.00%), evaluating (0.50%), and producing (0.00%) are very deficient. This disparity signifies an absence of complexity in the assessment items, potentially constraining students' opportunities to cultivate and exhibit critical thinking, problem-solving, and creativity. This distribution undermines the objective of promoting holistic cognitive development and indicates a necessity to realign examinations with Bloom's Taxonomy by incorporating additional higher-order activities.

The findings of the study highlights a concerning trend in educational assessments across various context, revealing an overemphasis on LOTS at the expense of HOTS. Evaluations of the test items for Grade 6 pupils across all quarters in San Mateo West Central School during the school year 2023-2024 shows a predominant focus on remembering, understanding, and applying level where it appeared to have the absence of evaluating, and creating questions which is also consistent on the study conducted by Utami *et al.* (2019). On the study Erdiana and Panjaitan (2023) found out that majority or 73.3% of the activities and test items in the classroom assessments focused on LOTS, while only 26.7% or very minimal addressed or focus on HOTS. Zamir *et al.*, (2023) identified a gap in language assessments, emphasizing the absence of analysis level questions and a mere .45% represents only an evaluation level items. These studies indicate that, despite the growing emphasis on 21st century skills, many test items still depend on assessing rote memorization and basic knowledge that limit students ability to develop critical thinking and problem solving skills. According to Zamani and Rezvani (2015) noted that even in advance academic programs, assessments still focused on LOTS questions rather than the essential complexities for deeper learning.

In a study conducted by Utami *et al.* (2019) highlighted the challenges faced by teachers in creating test items based on HOTS, pointing out that many teachers mistakenly categorize lower- order thinking questions as higher- order due to a lack of adequate training in Test construction. The findings stress the necessity of curricular enhancements and targeted teacher development programs. The integration of HOTS test items into classroom assessments is essential for improving cognitive rigor and promoting analytical, and creative thinking among students.

4.8. Insufficient representation of creating and evaluating questions

One of the weaknesses of the revised TOS is the absence of creating level questions across all quarters and the minimal occurrence of evaluating level items. Activities that encourage the creation of innovation, new ideas, and the synthesis of information are essential for enhancing one of the 21st century skill in education which is creativity. Similarly, assessing questions that evaluate judgement, reasoning and decision making is crucial for promoting critical thinking. The lack of evaluating and creating questions can be affected by the assessment design, limiting the test's ability to evaluate higher-order cognitive processes.

After the taxonomic content analysis, it shows that there is deficiency or small percentage of higher- order thinking skills test items. This aligns with Utami *et al.*, (2019) where half of their assessments reported to be under LOTS level while Abosalem (2016) found that majority of the assessed test items are under application and a quarter of it are under understanding level. Both studies revealed a concerning absence of questions that assesses HOTS.

Despite these limitations, Scully (2017) asserts that the incorporation of HOTS nto assessments is achievable through intentional design and pedagogical collaboration and with the proper use of table of specification.

4.9. Proposed course of action to enhance teachers test construction abilities

Assessment is a crucial component of teaching. Ensuring reliability, and validity in assessments is necessary to evaluate the extent of learned and acquired knowledge of students that is align with a well- defined TOS (Cañeda, 2024). Reliability guarantees that results remain consistent over time and among different learners, whereas validity confirsts that test items accurately measure what intends to measure (Krieglstein *et al.*, 2023). The Bloom's revised taxonomy posits that a well-structured TOS serves as a framework ensuring that test items are distributed across various cognitive levels with a balance representation of LOTS and HOTS (SRJ, 2021).

Karaman (2024) states that failure to adhere with the principles of test construction and TOS preparation may result to inconsistencies of test scores and affect the test item's reliability. Quality test items can enhance student learning but requires teachers' skill in developing reliable and valid test item with the help of a training and professional development program. One of the salient findings of the study is on the teacher's ability to construct test items particularly in aligning test items with the proper cognitive level under revised Bloom's taxonomy. The misclassification of cognitive levels and overemphasis of LOTS among test items of teachers can be resolve through strengthening teachers understanding or knowledge on test construction specifically in aligning test items in the cognitive level.

Hence, this study proposed "*Strengthening Teachers' Competence in Test Construction: A Capacity Development Program for Teachers*" as a targeted initiative aimed at enhancing teachers' ability to design valid, reliable, and pedagogically sound test items. This intervention aims to address the weaknesses of teachers understanding on the revised Bloom's taxonomy, test construction principles and the proper alignment of test items. This intervention is a combination of a training sessions, practical exercises and workshops to enhance teachers' ability in creating balance assessment incorporating both LOTS and HOTS questions which will emphasize the crucial role of creating test items that are valid and reliable.

5. CONCLUSION

Based on the summary presented, the following conclusions were drawn:

i. The predominance of lower-order thinking skills

(specifically understanding) in test items reflects a theoretical misalignment with Bloom's Revised Taxonomy, which advocates for a balanced development of cognitive processes. This suggests that assessment practices remain heavily content-recall oriented, thereby limiting opportunities for learners to demonstrate critical thinking, problem-solving, and knowledge transfer core goals of outcome-based education.

ii. The dominance of the applying level in the revised TOS indicates partial progress toward higher cognitive engagement but also reveals that teachers often misjudge the complexity of their test items. Theoretically, this misclassification undermines the validity of assessments by conflating procedural fluency with conceptual mastery. Practically, this highlights the need for assessment calibration and teacher upskilling in cognitive classification.

iii. The identified discrepancies between intended learning outcomes and actual test content point to fundamental weaknesses in test validity and reliability. From an assessment theory perspective, this inconsistency disrupts the construct validity of tests, as they fail to measure the full range of competencies outlined in the curriculum. Therefore, systematic alignment of test items with the Table of Specifications (TOS) and Bloom's cognitive domains is necessary to ensure assessment fidelity.

iv. Although the TOS ensures content coverage, its uneven cognitive distribution and misclassification of test items reveal structural gaps in teachers' assessment literacy. Theoretically, this reflects a lack of coherence between formative and summative assessment design principles. Practically, revising test construction protocols, integrating peer-review mechanisms, and institutionalizing assessment audits will improve test fairness and accuracy.

v. The findings collectively underscore the need for a program focused on strengthening teachers' test construction abilities. The proposed program "*Strengthening Teachers' Competence in Test Construction*" should incorporate theoretical sessions on assessment principles, practical workshops on cognitive alignment, peer evaluation of test items, and continuous feedback cycles. Such an initiative operationalizes assessment theory by bridging the gap between teachers' conceptual understanding and their practical application, thereby enhancing the validity, reliability, and equity of classroom assessments.

REFERENCES

Abate, T., & Mishore, E. (2024). Alignment analysis between teacher-made tests with the learning objectives in a selected school of central regional state of Ethiopia. *Heliyon, 10*(11).

Abosalem, Y. (2016). *Assessment Techniques and Students' Higher-Order Thinking Skills.*

AERA, APA, & NCME. (2018). *Standards for educational and psychological testing*. American Educational Research Association. https://doi.org/10.3102/9780935302611

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

Arsyad Arrafii, M. (2023). Assessment reform in Indonesia: Contextual barriers and opportunities for implementation. *Asia Pacific Journal of Education, 43*(1), 79-94.

Arta, G. Y. (2024). Assessment in education: Concepts, approaches, principles, types, and functions [Asesmen dalam pendidikan: Konsep, pendekatan, prinsip, jenis, dan fungsi]. *Jurnal Pendidikan, Bahasa Dan Budaya, 3*(3), 170-190.

Atna, Ari, & Ilalahi (2022). In An English Textbook For Year X Senior High School Students Based On Bloom's Taxomony.

Baddane, K., & Ennam, A. (2024). Measuring pedagogical transformation: A quantitative analysis of critical thinking integration in literary criticism for heightened student engagement and learning outcomes. *International Journal of Linguistics, Literature and Translation, 7*(1), 39-50.

Baird, J.-A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2019). Assessment and learning: Fields apart? Assessment in Education: Principles, Policy & Practice, 26(5), 515-537. https://doi.org/10.1080/0969594X.2019.1593103

Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford Publications. https://doi.org/10.1037/0000177-000

Barber, W., King, S., & Buchanan, S. (2015). Problem Based Learning and Authentic Assessment in Digital Pedagogy: Embracing the Role of Collaborative Communities. *Electronic Journal of e-Learning, 13*, 59-67.

Başol, G., Balgalmiş, E., Karlı, M.G., & Öz, F.B. (2016). Content analysis of TEOG mathematics items based on MONE attainments, TIMSS levels, and reformed Bloom Taxonomy TEOG sınavı matematik sorularının MEB kazanımlarına, TIMSS seviyelerine ve yenilenen Bloom Taksonomisine göre incelenmesi. *Journal of new results in science, 13*, 5945-5967.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. Longman.

Brookhart, S. M. (2018). *How to create and use rubrics for formative assessment and grading*. ASCD. https://www.ascd.org/books/how-to-create-and-use-rubrics-for-formative-assessment-and-grading

Brookhart, S. M. (2018). *How to Design Questions and Tasks to Assess Student Thinking*. ASCD.

Brookhart, S. M. (2020). *Assessment and grading in classrooms*. ASCD. https://www.ascd.org/books/assessment-and-grading-in-classrooms

Cañeda, M. (2024). *Ensuring validity and reliability in Algebra midterm assessment: A comprehensive approach to test development and analysis.*

Cañeda, M. E., Logroño, J. J. F., & Culibra, C. D. (2024). Test questionnaire development on selected topics in calculus 1. *Ignatian International Journal for Multidisciplinary Research, 2*(8), 1363-1376.

Chew, S. L., & Cerbin, W. J. (2021). The cognitive challenges of effective teaching. *The Journal of Economic Education, 52*(1), 17-40.

Choy, S.C., Abdul, T.A., & Cheah, P.K. (2009). Teacher Perceptions of Critical Thinking Among Students and its Influence on Higher Education. *The International Journal of Teaching and Learning in Higher Education, 20*, 198-206.

Dacles, G. G. (2024). 21st-Century Teaching Competencies, Learning Skills and Science Content Knowledge of STEM Students Among the DepEd Schools in the Province of Capiz: A Basis for Enhancement Program. *International Journal of Multidisciplinary: Applied Business and Education Research, 5*(12), 5006-5039.

Dagostino, L., Carifio, J., Bauer, J., Zhao, Q., & Hashim, N.H. (2014). Assessment of a Reading Comprehension Instrument as It Relates to Cognitive Abilities as Defined by Bloom's Revised Taxonomy. *Current Issues in Education, 17*.

de Sousa Oliveira, M. P. (2023). Procedural/Rote Learning Vs Meaningful Learning: Which Do First Year Undergraduate Students Prefer?. In *Edulearn23 Proceedings* (pp. 1681-1690). IATED.

DepEd Order no. 031, s. (2021). Interim Guidelines for Assessment and Grading in Light of the Basic Education Learning Continuity Plan.

DepEd Order no. 29, s. (2017). Policy Guidelines on System Assessment in the K to 12 Basic Education Program.

Djamalovna, D. M. (2024). Methods and Tools for Assessing Student Competencies. *Current Research Journal of Philological Sciences, 5*(10), 19-24.

Driana, E., & Ernawati, E. (2019). Teachers'understanding And Practices In Assessing Higher Order Thinking Skills At Primary Schools. *Acitya: Journal of Teaching and Education, 1*(2), 110-118.

Driana, E., & Ernawati, E. (2024). Conception and Implementation of HOTS-Based Reading Literacy Assessment in English [Konsepsi dan Implementasi Penilaian Literasi Membaca Berbasis HOTS dalam Bahasa Inggris]. *Jurnal Manajemen Pendidikan dan Ilmu Sosial (JMPIS), 5*(2).

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking Skills and Creativity, 12*, 43–52. https://doi.org/10.1016/j.tsc.2013.12.004

Epinasti, G., Y. Slamet, S., & Yamtinah, S. (2021). Teacher's Perception towards Assessing of Higher Order Thinking Skills (HOTS) in Elementary Schools. *Proceedings of the 5th International Conference on Learning Innovation and Quality Education*.

Erdiana, N., & Panjaitan, S. (2023). How is HOTS integrated into the Indonesian high school English textbook?. *Studies in English Language and Education, 10*(1), 60-77.

Eswaran, U. (2024). Project-based learning: Fostering collaboration, creativity, and critical thinking. In *Enhancing education with intelligent systems and data-driven instruction* (pp. 23-43). IGI Global.

Faradella, N. E. (2024). *Teachers'strategies And Challenges In Conducting Higher Order Thinking Skills (Hots)–Based Activities In English Language Classroom* (Doctoral dissertation, Universitas PGRI Semarang).

Febriyana, F., & Harjanto, I. (2023). Cognitive levels of questions by Indonesian teachers of English. *Journal of English Language Teaching and Linguistics, 8*(2), 2023.

Febriyana, R., & Harjanto, I. (2023). The analysis of HOTS and LOTS test items in teacher-made assessments. *International Journal of Educational Research & Innovation, 18*(1), 35-50. https://doi.org/10.1080/00220671.2023.2084956

Forehand, M. (2010). Bloom's taxonomy: From knowledge to metacognition. In M. Orey (Ed.), *Emerging perspectives on learning, teaching, and technology*. The University of Georgia.

Gareis, C. R., & Grant, L. W. (2020). *Teacher-made assessments: How to connect curriculum, instruction, and student learning*. Routledge. https://doi.org/10.4324/9781003013836

Ginting, P., Hasnah, Y., Hasibuan, S. H., & Batubara, I. H. (2021). Evaluating cognitive level of final semester examination questions based on Bloom's revised taxonomy. *Al-Ishlah: Jurnal Pendidikan, 13*(1), 186-195.

Ginting, P., Suryani, N., & Herman, J. (2021). The implementation of HOTS-oriented assessment in university-level examinations. *Journal of Educational Assessment, 45*(3), 215-230. https://doi.org/10.1080/02602938.2021.2001538

Gotch, C. M., & French, B. F. (2020). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice, 39*(2), 14-24. https://doi.org/10.1111/emip.12350

Gunawan, I., Rahmat, A., & Suryani, A. (2021). Higher-order thinking skills in mathematics assessment: A systematic review. *Journal of Mathematics Education Research, 14*(2), 78-94. https://doi.org/10.1016/j.jmer.2021.002

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.

Herman, J. L., & Linn, R. L. (2020). On the road to assessing deeper learning: The status of Smarter Balanced and PARCC

assessment consortia. *Educational Assessment, 25*(4), 245-263. https://doi.org/10.1080/10627197.2020.1836013

Huang, Y. M., Silitonga, L. M., & Wu, T. T. (2022). Applying a business simulation game in a flipped classroom to enhance engagement, learning achievement, and higher-order thinking skills. *Computers & Education, 183*, 104494.

Johnson, R. L., Penny, J., & Gordon, B. (2021). *Assessing performance: Designing, scoring, and validating performance tasks.* Guilford Publications. https://doi.org/10.4324/9781315691136

Karaman, M. R. (2024). Are Lesson Plans Created by ChatGPT More Effective? An Experimental Study. *International Journal of Technology in Education, 7*(1), 107-127.

Kim, M. K., Patel, R. A., Uchizono, J. A., & Beck, L. (2012). Incorporation of Bloom's taxonomy into multiple-choice examination questions for a pharmacotherapeutics course. *American journal of pharmaceutical education, 76*(6), 114.

Kotsis, K. T. (2024). Integrating Inquiry-based Learning in the New Greek Primary Science Curriculum. *European Journal of Education and Pedagogy, 5*(6), 63-71.

Kouicem, K., & BOULAKHAL, K. (2025). Identifying the Cognitive Complexity of Teacher-Made Examination Questions through the Lens of Bloom's Taxonomy. *ATRAS journal, 6*(1), 284-300.

Krieglstein, F., Beege, M., Rey, G. D., Sanchez-Stockhammer, C., & Schneider, S. (2023). Development and validation of a theory-based questionnaire to measure different types of cognitive load. *Educational Psychology Review, 35*(1), 9.

Maryamah, S. E., Wati, S., & Suharto, P. P. (2024). Analyzing Higher Order Thinking Skills in Reading Exercises of Indonesian EFL Textbook. *SALEE: Study of Applied Linguistics and English Education, 5*(2), 560-570.

Miterianifa, M., Ashadi, A., Saputro, S., & Suciati, S. (2021). Higher Order Thinking Skills in the 21st Century: Critical Thinking. *Proceedings of the 1st International Conference on Social Science, Humanities, Education and Society Development, ICONS,* 2020, 30 November, Tegal, Indonesia.

Musliha, S., Sudana, D., & Wirza, Y. (2021, November). The Analysis of Higher Order Thinking Skills (HOTs) in the test questions constructed by English teachers. In *Fifth International Conference on Language, Literature, Culture, and Education (ICOLLITE 2021)* (pp. 610-617). Atlantis Press.

Najuah, N., Rangkuti, S. M., Sidiq, R., Lukitoyo, P. S., & Diansyah, A. (2024). LOTS vs HOTS: Evaluation of History Textbooks for Class XII Senior High School. *AL-ISHLAH: Jurnal Pendidikan, 16*(2), 737-747.

Newton, P. E., & Shaw, S. D. (2018). *Validity in educational and psychological assessment.* SAGE Publications. https://doi.org/10.4135/9781473920453

Orongan, R. C. (2020). Reliability Analysis on Teachers' Quarterly Classroom Assessment in Basic Education. *Liceo Journal of Higher Education Research, 16*(1).

Pellegrino, J. W. (2022). The challenge of aligning classroom assessment with contemporary learning theory. *Educational Psychologist, 57*(3), 179-194. https://doi.org/10.1080/00461520.2022.2036504

Pratama, A. H., & Retnawati, H. (2018). Urgency of Higher Order Thinking Skills (HOTS) content analysis in mathematics textbook. *Journal of Physics: Conference Series, 1097*(1), 012147. https://doi.org/10.1088/1742-6596/1097/1/012147

Prihastuti, I., & Widodo, A. (2019, February). Teachers understanding about cognitive level on science test items. In *Journal of Physics: Conference Series* (Vol. 1157, No. 2, p. 022054). IOP Publishing.

Prihastuti, I., & Widodo, A. (2020, March). Cognitive level analysis of science item tests on secondary school assessment. *In Journal of Physics: Conference Series, 1521*, (4), 042106). IOP Publishing.

Putra, R. P., Surya, E., & Ahmad, N. (2020). The importance of higher-order thinking skills (HOTS) in mathematics learning. *International Journal of Learning and Teaching, 12*(1), 33-40. https://doi.org/10.2991/ijlt-20.2020.05

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*, 744-758.

Rania, A. (2023). *Investigating the Effectiveness of Using Bloom's Taxonomy to Develop EFL Learners High Order Thinking Skills in Formative Assessment The Case of First year students of English Kasdi Merbah University-Ouargla* (Doctoral dissertation, Kasdi Merbah Ouargla University).

Retnawati, H., Djidu, H., Kartianom, Anazifa, R. D., & Usodo, B. (2018). Teachers' knowledge about higher-order thinking skills and its learning strategy. *Problem of Education in the 21st Century, 76*(2), 215-230. https://doi.org/10.33225/pec/18.76.215

Richter-Beuschel, L., Grass, I., & Bögeholz, S. (2018). *How to Measure Procedural Knowledge for Solving Biodiversity and Climate Change Challenges.* Education Sciences.

Saputra, A., Purnama, W., & Hidayah, S. (2022). Implementing Bloom's Revised Taxonomy in assessing students' cognitive development. *Educational Assessment Journal, 17*(3), 112-126. https://doi.org/10.3102/eaj.2022.017

Saputra, R. P. S., & Fatimah, S. (2022). The implementation of Revised Bloom's Taxonomy in online assessment: An analysis of cognitive level distribution. *Journal of Education and Learning Studies, 5*(1), 76-84. https://doi.org/10.32996/jels.2022.5.1.9

Sari, N. V. N., & Abdurrahman, S. (2019). Developing and validating of the three tier diagnostic test based'higher order thinking skills' instrument. *Dinamika Jurnal Ilmiah Pendidikan Dasar, 11*(2), 87-93.

Schmoker, M. (2018). *Focus: Elevating the essentials to radically improve student learning*. Ascd.

Sebastian, M. (2020). The classification dilemma: How pre-service mathematics teachers perceive cognitive complexity. *Journal of Educational Psychology, 112*(6), 1082-1095. https://doi.org/10.1037/edu0000412

Sebastian, M. A. (2020). Classification of Test Items Written by Mathematics Pre-Service Teachers. *International Electronic Journal of Mathematics Education, 15*(2).

Setiawan, H., Wulandari, S., & Firdaus, F. (2019). Evaluating cognitive levels in test construction: The dominance of LOTS over HOTS. *Indonesian Journal of Education, 25*(4), 101-117. https://doi.org/10.21093/ije.2019.25

Setiawati, I., & Corebima, A. D. (2018). The correlation between critical thinking and the HOTS of students in biology learning. *Journal of Baltic Science Education, 17*(6), 1055-1066. https://doi.org/10.33225/jbse/18.17.1055

Sohdi, R. (2025). A design-based approach to bridging the gap between university learning and classroom pedagogy in Initial Teacher Education. *Journal of Learning Development in Higher Education, 34*.

Srinivasa, K. G., Kurni, M., & Saritha, K. (2022). *Learning, Teaching, and Assessment Methods for Contemporary Learners* (pp. 1-15). Singapore: Springer.

Suryani, A., Hadi, R., & Yusuf, M. (2020). Exploring teachers' perspectives on Bloom's Revised Taxonomy in curriculum development. *Asian Journal of Educational Research, 5*(2), 88-99. https://doi.org/10.5281/zenodo.3987212

Suskie, L. (2018). *Assessing student learning: A common sense guide*. John Wiley & Sons.

Tarbiyah, F.I., Keguruan, D.I., Ar-Raniry-Banda, U., & Aceh (2024). *The Power of Evaluation: Transforming Teaching and Learning*. Jurnal Ilmiah Mahasiswa.

Thampi, K. P. P. S., & Yongsorn, C. (2023). *Developing Critical Thinking Skills And Dispositions Among Higher Education Students In Thailand* (Doctoral dissertation, Srinakharinwirot University).

Utami, F. D., Hidayati, D., & Anisah, N. (2019). The dominance of LOTS in English test items: A case study in Indonesian high schools. *Indonesian Journal of Applied Linguistics, 9*(2), 456-470. https://doi.org/10.17509/ijal.v9i2.16351

Utami, F. D., Nurkamto, J., & Marmanto, S. (2019). Higher-order thinking skills on test items designed by English teachers: A content analysis. *International Journal of Educational Research Review, 4*, 756-765.

Valentiyo, R., Hidayah, J., & Edy, S. (2024). *Higher Order Thinking Skill Questions In Reading Comprehension Test Made By English Teacher Of Junior High School Based On The Implementation Of Merdeka Curriculum* (A Descriptive Research On Smpn 01 Rejang Lebong) (Doctoral dissertation, Institut Agama Islam Negeri Curup).

Wibowo, S. A., Nuro, F. R. M. A., & Lestari, D. N. (2024). Project-based learning with interactive video media to improve students' HOTS. *Jurnal Bidang Pendidikan Dasar, 8*(2), 105-113.

Widana, I. W. (2017). Higher Order Thinking Skills Assessment (HOTS). *JISAE: Journal of Indonesian Student Assessment and Evaluation, 3*(1), 32–44. https://doi.org/10.21009/jisae.v3i1.4859

Widana, I. W. (2020). The impact of higher-order thinking skills assessment on students' critical thinking skills. *Journal of Education Research and Evaluation, 4*(1), 31-38. https://doi.org/10.23887/jere.v4i1.23909

Widana, I. W., Dewi, A., & Suastra, I. W. (2019). Higher-order thinking skills assessment as an innovative learning model. *International Journal of Innovation and Learning, 26*(1), 45-57. https://doi.org/10.1504/IJIL.2019.100227

Xu, Y., & Brown, G. T. L. (2018). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education, 70*, 1-11. https://doi.org/10.1016/j.tate.2017.11.007

Zamani, G., & Rezvani, R. (2015). 'HOTS'in Iran's Official Textbooks: Implications for Material Design and Student Learning. *Journal of Applied Linguistics and Language Research, 2*(5), 138-151.

Zamir, S., Jan, H., Ansari, S. R., Hussain, R., & Rehman, M. U. (2023). Analyzing Language Assessment from the Cognitive Lens. *Academy of Education and Social Sciences Review, 3*(4), 440-452.

Zana, F. M., Sa'dijah, C., Susiswo, L. A., & Zulnaidi, H. (2024). Curriculum and teacher assessment practices in mathematics learning: Alignment with higher order thinking skills in Indonesian secondary schools. *Journal on Mathematics Education, 15*(4), 1311-1334.

Zhou, M. (2023). Significance of assessment in learning: The role of educational assessment tools. *Science Insights Education Frontiers, 18*(2), 2881-2883. https://doi.org/10.15354/sief.23.co215

Zoller, U., & Pushkin, D. (2007). Matching Higher-Order Cognitive Skills (HOCS) promotion goals with problem-based laboratory practice in a freshman organic chemistry course. *Chemistry Education Research and Practice, 8*, 153-171. https://doi.org/10.1039/B6RP90028C

Zoller, U., & Tsaparlis, G. (1997). Higher and lower-order cognitive skills: The case of chemistry. *Research in Science Education, 27*, 117–130. https://doi.org/10.1007/BF02463036