

Evaluating Teacher-Made Test in STEM, Math and Science: Alignment, Cognitive Taxonomic Levels, and Item Construction

Glanda P. Garcia

Doctor of Philosophy in Science Education, Major in Mathematics, Nueva Vizcaya State University, Bayombong, Nueva Vizcaya, Philippines

Email of Corresponding Author: glenlp.1234@gmail.com

Publication history:

Received: September 19, 2026

Accepted: October 20, 2026

Revised: October 17, 2026

Published: October 21, 2026

International Journal of Sustainable and Integrated Studies, Vol. 1, No. 3, pp. 695–701, October 2026.

Abstract

This study evaluated the quality of teacher-made examinations in Science, Technology, Engineering, and Mathematics (STEM), Science, and Mathematics at Bambang National High School, Nueva Vizcaya, Philippines, focusing on curriculum alignment, cognitive demands, and test construction quality. Guided by Bloom's Revised Taxonomy and curriculum–assessment alignment principles, the study employed a descriptive research design using qualitative content analysis of 18 second-quarter examination papers and their corresponding Tables of Specifications (TOS) from School Year 2024–2025. Documents were selected through purposive and convenience sampling and evaluated using structured content analysis protocols, curriculum competency mapping, and a test construction observation guide. Frequencies and percentages were used to summarize the findings. Results revealed varying levels of assessment alignment, with Physics recording the highest examination–TOS alignment (96.67%) and Empowerment Technology the lowest (48.33%). Curriculum competency alignment was generally high, although Grade 8 Mathematics recorded only 31.58%. Remembering-level questions predominated in Science and Earth and Life Science, while Mathematics and Physics demonstrated broader cognitive distributions. Common construction flaws included implausible distractors, multiple correct answers, overlapping choices, and negatively worded items. The study concluded that assessment quality requires improvements in curriculum alignment, cognitive balance, and item construction. Regular assessment reviews, teacher professional development, collaborative planning, and item validation were recommended. The study supports Sustainable Development Goal 4 (Quality Education) by promoting curriculum-aligned and cognitively appropriate classroom assessments. Its institutional sustainability contribution lies in providing evidence to strengthen teacher assessment competencies, systematic examination review, and continuous improvement of educational assessment practices.

Keywords: Assessment Quality; Bloom's Revised Taxonomy; Curriculum Alignment; STEM Education; Tables of Specifications; Teacher-Made Tests; Test Construction Quality

1. Introduction

Teacher-made tests are designed to measure learning outcomes based on instructional objectives and classroom content. When properly constructed, they support knowledge retention and meaningful learning (Murphy et al., 2023). Their effectiveness depends on content validity, appropriate cognitive demands, and alignment with curriculum standards (Areekkuzhiyil, 2021; Thompson, 2008; Martone & Sireci, 2009; Tiana & Maruf, 2024). Bloom's Taxonomy and test blueprints guide the development of assessments that measure different cognitive processes while maintaining clarity and consistency (Ajayi, 2024; Lemons & Lemons, 2013; Chiavaroli, 2017).

The study was guided by Bloom's Taxonomy and curriculum–assessment alignment principles. Bloom's Taxonomy classifies cognitive processes from remembering to creating, providing a basis for evaluating the cognitive demands of examination items (Ajayi, 2024; Lemons & Lemons, 2013). Curriculum alignment emphasizes consistency among learning competencies, TOS, and assessment items to strengthen content validity (Martone & Sireci, 2009; Tiana & Maruf, 2024). The conceptual framework followed an input–process–output model. Inputs included teacher-made tests, TOS, curriculum guides, and taxonomic classifications. Document analysis involved comparing examinations with their TOS, examining curriculum alignment, classifying cognitive levels, and identifying item construction flaws. Outputs consisted of alignment frequencies, cognitive classifications, and identified assessment weaknesses.

Previous studies have identified weaknesses in curriculum alignment, cognitive classification, and test construction (Ajogbeje, 2023; Thompson, 2008). However, localized evidence concerning these assessment dimensions at Bambang National High School remained limited. This study evaluated second-quarter teacher-made examinations during School Year 2024–2025, covering Mathematics and Science in Grades 7–10 and selected STEM-related subjects in Grades 11–12. Specifically, it determined the alignment between examinations and TOS according to taxonomic levels, the alignment between TOS and curriculum competencies, the cognitive classifications of test items, and common construction flaws. The investigation was limited to the collected second-quarter examinations and corresponding TOS.

The study aligns with Sustainable Development Goal 4 (Quality Education) by supporting improved classroom assessment quality and curriculum alignment. Evaluating examination content, cognitive demands, and construction flaws can provide evidence for strengthening instructional practices and teacher assessment literacy. Effective assessment and professional development support meaningful learning and educational improvement (Adarkwah, 2021; De Vries et al., 2022).

The study contributes to educational and institutional sustainability by providing evidence for improving assessment standards, teacher competencies, and curriculum implementation. Its findings may guide teachers in constructing valid examinations, assist head teachers in monitoring assessment practices, and support school administrators in developing evidence-based assessment policies. By promoting systematic examination review and continuous instructional improvement, the study supports sustainable educational quality in STEM, Science, and Mathematics.

2. Material and methods

2.1 Research Design

The study utilized a descriptive research design employing qualitative content analysis to systematically evaluate teacher-made examinations. The analysis focused on the alignment of examination items with their corresponding Tables of Specifications (TOS) and curriculum competencies, classification of cognitive demands using Bloom's Revised Taxonomy, and identification of common test construction flaws.

2.2 Locale of the Study

The study was conducted at Bambang National High School, a public secondary school in Bambang, Nueva Vizcaya, Philippines, during School Year 2024–2025. The institution served approximately 3,801 students, comprising 2,254 Junior High School and 1,547 Senior High School students, with 147 teaching personnel and 18 non-teaching staff. The investigation specifically focused on the STEM, Mathematics, and Science department.

2.3 Respondents of the Study

The study examined 18 teacher-made second-quarter examination papers and their corresponding TOS during the first semester of School Year 2024–2025. These documents comprised four Mathematics examinations from Grades 7–10, four Science examinations from Grades 7–10, and ten STEM-related examinations. The STEM subjects included General Mathematics (1), Pre-Calculus (1), Earth and Life Science (4), Physics (1), and Empowerment Technology (3). These examination documents served as the primary data sources for evaluating curriculum alignment, cognitive demand, and construction quality.

2.4 Sample and Sampling Procedure

Purposive and convenience sampling techniques were employed to select 18 available teacher-made examinations and their corresponding TOS. Purposive sampling ensured the inclusion of documents relevant to STEM, Science, and Mathematics, while convenience sampling limited the selection to examination materials accessible from the school's archived records. The selection focused exclusively on available second-quarter examination documents.

2.5 Research Instrument

Structured content analysis protocols were used to evaluate curriculum alignment, cognitive demand, and test construction quality. Curriculum alignment was assessed by mapping examination items against their

corresponding TOS and comparing TOS content with the learning competencies prescribed in the official K to 12 Curriculum Guides. Cognitive classification followed Bloom's Revised Taxonomy, categorizing examination items into remembering, understanding, applying, analyzing, evaluating, and creating based on cognitive verbs and question stems. Test construction quality was evaluated using an observation guide informed by Module 9, Principles of Test Construction, discussed by Dr. Jane Navalta. The evaluation identified ambiguous wording, answer clues, inappropriate difficulty, irrelevant information, negative phrasing, double-barreled questions, unclear instructions, implausible distractors, and leading or biased wording.

2.6 Data Gathering Procedure

Permission to access examination materials was obtained from the school principal and the head teacher of the STEM, Science, and Mathematics department. Following approval, previously submitted second-quarter examination papers and their corresponding TOS were retrieved from the head teacher's office. Official curriculum guides were downloaded from the Department of Education portal. The researcher systematically mapped the TOS against prescribed learning competencies and compared examination items with their corresponding TOS to determine content and cognitive alignment. Each item was subsequently classified according to Bloom's Revised Taxonomy and examined for construction flaws using the established evaluation protocols.

2.7 Data Analysis Techniques

Content mapping and descriptive statistics were employed to analyze the examination documents. Alignment between teacher-made tests and their corresponding TOS was evaluated by comparing content coverage, item distribution, and cognitive classifications. The TOS were also examined against official curriculum guides to determine alignment with prescribed learning competencies. Frequency counts and percentages were used to summarize curriculum alignment and the distribution of examination items across the six cognitive levels of Bloom's Revised Taxonomy. Identified construction flaws were categorized and quantified by subject area to determine recurring weaknesses in examination development. These procedures provided a descriptive evaluation of assessment alignment, cognitive complexity, and test construction quality across the selected subjects.

3. Results and Discussion

3.1 Alignment of Teacher-Made Tests with Assessment Specifications and Curriculum Competencies

3.1.1 Alignment Between Teacher-Made Tests and Tables of Specifications

The alignment between teacher-made tests and their corresponding TOS varied across subjects. Physics recorded the highest alignment at 96.67% (58 items), followed by Grade 7 Mathematics at 90.00% (54 items). Mathematics subjects demonstrated alignment ranging from 66.67% to 90.00%, while Science subjects ranged from 73.33% to 83.33%. General Mathematics and Pre-Calculus recorded 80.00% and 81.67%, respectively. Lower alignment was observed in Earth and Life Science, ranging from 58.33% to 68.33%, and Empowerment Technology, ranging from 48.33% to 63.33%. Empowerment Technology (T3) recorded the lowest alignment at 48.33% (29 items). These findings indicate varying consistency between intended and actual cognitive demands across subject areas. Higher alignment suggests closer adherence to assessment specifications, whereas lower alignment indicates weaknesses in cognitive classification and test planning. Khan and Aslam (2024) emphasized the importance of systematic assessment blueprinting, while Galle et al. (2024) and Oke James Ajogbeje (2023) identified insufficient familiarity with TOS as a factor contributing to assessment misalignment. Similar inconsistencies were reported by Tiana and Maruf (2024) and Abate and Mishore (2024), reinforcing the need for improved alignment between examinations and intended learning objectives.

3.1.2 Alignment Between Assessment Specifications and Curriculum Learning Competencies

Curriculum competency alignment was generally high but varied among subjects. Grade 7 Mathematics, Grades 7–10 Science, all four Earth and Life Science examinations, Empowerment Technology (T1), Physics, and Pre-Calculus recorded 100% alignment. General Mathematics achieved 78.57%, while Grade 8 Mathematics recorded the lowest alignment at 31.58%, with 14 first-quarter learning competencies included. Grade 9 Mathematics achieved 50.00%, Grade 10 Mathematics 57.14%, and Empowerment Technology (T2) 80.00%. Empowerment Technology (T3) recorded 50.00% alignment, with four competencies absent from the official curriculum guide. Grade 9 Mathematics and Pre-Calculus also included two and three first-quarter competencies, respectively. These discrepancies indicate differences in the implementation of prescribed curriculum competencies during assessment development. The inclusion of competencies from other grading periods or outside the official curriculum may weaken content validity and the accuracy of learning outcome measurement. Irmala Sukendra (n.d.) similarly identified mismatches between teacher-made examinations and instructional objectives, while Ahmed et al. (2025) emphasized the importance of alignment with prescribed science learning outcomes. Banjal et al. (2025) further highlighted curriculum alignment challenges in Philippine education, supporting the need for consistent adherence to curriculum guides and assessment specifications.

3.2 Cognitive Taxonomic Classification of Teacher-Made Test Items

3.2.1 Distribution of Lower-Order and Higher-Order Cognitive Levels

The cognitive classification revealed a predominance of remembering-level questions, particularly in Science and Earth and Life Science. Grade 8 Science recorded the highest proportion of remembering items at 85.00%, followed by Earth and Life Science (T4) at 81.67% and Earth and Life Science (T3) at 71.67%. In contrast, Grade 8 Mathematics and Grade 10 Mathematics emphasized applying-level questions at 58.33% and 56.67%, respectively. General Mathematics contained 51.66% remembering and 43.33% applying items. Physics demonstrated a broader cognitive distribution, comprising 46.67% remembering, 18.33% understanding, 18.33% applying, 13.33% analyzing, and 3.33% evaluating. Pre-Calculus included 55.00% remembering, 25.00% applying, and 10.00% analyzing items. No creating-level items were reported. The predominance of remembering-level questions suggests that many examinations emphasized knowledge recall rather than complex cognitive processes. Although Mathematics and Physics demonstrated greater representation of application and analytical skills, higher-order thinking remained limited across several subjects. Kini et al. (2025) similarly identified the predominance of lower-order questions in teacher-developed assessments, while Orhani's (2024) findings emphasized Bloom's Taxonomy as a framework for improving cognitive assessment quality. Stringer et al. (2021) highlighted discrepancies between intended and perceived cognitive demands, whereas Colak Seker and Demir's (2022) findings demonstrated similar limitations in science assessment questions. Raymundo's (2021) research further supported the need for balanced cognitive demands in Philippine mathematics examinations.

3.3 Common Flaws in Teacher-Made Test Item Construction

3.3.1 Frequency and Types of Test Construction Flaws

The evaluation identified several construction flaws across subject areas, with the highest concentrations observed in Earth and Life Science and Physics. These included 84 instances of implausible distractors, 60 items with more than one correct answer, 18 negatively stated items, and 14 instances of overlapping choices or grammatical cues. Empowerment Technology exhibited eight instances of unclear or complex sentence structures and seven negatively stated items. Mathematics recorded 12 instances of multiple correct answers and 11 implausible distractors, while Science contained eight implausible distractors and six incomplete stems. General Mathematics and Pre-Calculus exhibited five instances of overlapping choices or grammatical cues, alongside fewer occurrences of other construction weaknesses. These findings indicate that deficiencies in item clarity, answer options, and question structure may compromise assessment validity and the accurate interpretation of student performance. Poorly constructed teacher-made assessments and inadequate test blueprinting can weaken content validity. Negatively worded questions may contribute to student misinterpretation and construct-irrelevant variance. The observed construction flaws underscore the importance of systematic item review and adherence to established assessment development principles.

3.4 Overall Assessment Quality Across STEM, Science, and Mathematics

The findings revealed variations in curriculum alignment, cognitive demands, and test construction quality across the examined subjects. Although several examinations demonstrated high alignment with their corresponding TOS and curriculum competencies, inconsistencies were evident in selected Mathematics, Earth and Life Science, and Empowerment Technology assessments. Most Science and Earth and Life Science examinations emphasized remembering-level questions, while Mathematics and Physics demonstrated greater representation of application and analytical skills. Construction flaws, particularly implausible distractors and multiple correct answers, were also identified across subject areas.

4. Conclusion & Recommendations

The study revealed varying levels of quality in teacher-made STEM, Science, and Mathematics examinations at Bambang National High School. While most subjects demonstrated high alignment with their corresponding Tables of Specifications (TOS) and curriculum competencies, inconsistencies were identified in Earth and Life Science, Empowerment Technology, and Grade 8 Mathematics. Many examinations, particularly in Science and Earth and Life Science, predominantly assessed lower-order thinking skills, whereas Mathematics and Physics demonstrated a broader distribution of cognitive demands. Common construction flaws included multiple correct answers, implausible distractors, overlapping choices, and negatively worded items. These findings highlight the need to strengthen curriculum alignment, cognitive assessment, and test construction practices to improve assessment validity and reliability.

It is recommended that school administrators implement regular assessment reviews and curriculum alignment audits to ensure consistency between teacher-made examinations, TOS, and prescribed learning competencies. Professional development programs should strengthen teachers' competencies in Bloom's Taxonomy, higher-order question construction, and test development principles, particularly in subjects exhibiting greater assessment weaknesses. Peer review, collaborative planning, and item validation should be encouraged before examination administration. Future evaluations should involve subject-area experts and examine additional assessment

components, including the alignment of TOS with daily learning logs and the formats of examination papers and test specifications.

The study was limited to 18 available teacher-made second-quarter examination papers and their corresponding TOS in selected STEM, Science, and Mathematics subjects at Bambang National High School during School Year 2024–2025. The documents were selected through purposive and convenience sampling, restricting the analysis to accessible assessment materials. The evaluation focused on curriculum alignment, cognitive taxonomic classifications, and item construction flaws, excluding examinations from other grading periods and subjects not represented in the collected documents. The assessment review also did not incorporate subject-area content experts, limiting the depth of discipline-specific evaluation. Consequently, the findings were confined to the examined documents and may not represent assessment practices across other subjects, grading periods, or educational institutions.

5. Compliance with ethical standards

The author acknowledges Bambang National High School, Nueva Vizcaya, Philippines, for granting institutional permission to access archived teacher-made examinations and their corresponding Tables of Specifications. The study employed non-invasive documentary analysis of existing educational records without direct recruitment of human participants, and confidentiality was maintained by excluding personally identifiable information from the reported findings. The authors' conflict-of-interest and funding declarations require confirmation prior to publication.

REFERENCES

- Abate, T., & Mishore, E. (n.d.). Alignment analysis between teacher-made tests with the learning objectives in a selected school of central regional state of Ethiopia. *Heliyon*. <https://www.cell.com/heliyon>
- Adarkwah, M. A. (2021). The power of assessment feedback in teaching and learning: A narrative review and synthesis of the literature. *SN Social Sciences*, 1(75). <https://doi.org/10.1007/s43545-021-00086-w>
- Agawin, M. B., & Alferez, R. C. (2024). Teachers challenges in the conduct of assessment and their teaching performance. *Psychology and Education: A Multidisciplinary Journal*, 26(6), 620–632. <https://www.google.com/search?q=https://doi.org/10.5281/zenodo.13931549>
- Ahmed, K., Nazish, I., Hussain, T., Abbas, Y., Haseeb, M., & Khan, S. (2025). Identifying gaps in science curriculum student learning outcomes and Grade V students' achievement. *Indus Journal of Social Sciences*, 3(2), 243–258. <https://doi.org/10.1234/ijss.v3i2.09876>
- Ajayi, J. (2024). Blooms taxonomy. Structural Optimization. <https://www.researchgate.net/publication/380814622>
- Ajogbeje, O. J. (2023). Ensuring the content validity of teacher-made tests through the use of a table of specifications. *Zenodo*. <https://doi.org/10.5281/zenodo.8405151>
- Aprillia, A., Qadar, R., & Efwinda, S. (2023). Using revised Bloom's taxonomy to evaluate the cognitive levels of questions in Indonesian high school physics textbooks. *International Journal of STEM Education for Sustainability*, 3(1), 195–211. <https://doi.org/10.52889/ijses.v3i1.93>
- Areekkuzhiyil, S. (2021). Issues and concerns in classroom assessment practices. *Edutracks*, 20(8), 20–23. <https://www.researchgate.net>
- Aşıkcın, M., & Uygun, N. (2025). A taxonomic analysis of the questions prepared by prospective primary teachers for primary school mathematics and Turkish language courses. *International Journal of Education & Literacy Studies*, 11(3), 286–296. <https://doi.org/10.7575/aiac.ijels.v.11n.3p.286>
- Banjál, E. B., Berame, F. J. V., & Elesio, J. M. (2025). Harmonizing education: A case study on the constructive alignment approach to crisis in teacher education curriculum and licensure examination competencies. *International Journal of Research in Social Science and Humanities*, 6(1), 24. <https://doi.org/10.47505/IJRSS.2025.1.3>
- Becevic, S. (2023). When teachers construct tests for assessing students' competencies: A taxonomy. *Educational Studies in Mathematics*, 114(2), 315–336. <https://doi.org/10.1007/s10649-023-10233-0>
- Cañeda, M. E., Amar, R. J., & Lucin, E. (2024). Development of test questionnaire on selected topics in Calculus 1 (final term). *International Journal of Research and Scientific Innovation*, 11(8), 20. <https://doi.org/10.51244/IJRSI.2024.1108020>
- Colak Seker, Z., & Demir, C. (2022). Investigation of the evaluation questions in 5th grade science textbooks according to revised Bloom's taxonomy. *The Turkish Online Journal of Educational Technology, Special Issue for IETC*, 142–149.

- Darmawan, M., Sudarsono, Riyanti, D., Sutapa, Y. G., & Sumarni. (2022). A test-items analysis of English teacher-made test. *Journal of English Education and Teaching (JEET)*, 6(4), 498–513. <https://doi.org/10.33369/jeet.6.4.498-513>
- Department of Education. (2015). DO s 2015 No. 8: Policy guidelines on classroom assessment for the K to 12 basic education programs. www.deped.gov.ph
- Department of Education. (2016). General shaping paper bureau of curriculum development. cd.csdd@deped.gov.ph
- De Vries, J. A., Dimosthenous, A., Schildkamp, K., & Visscher, A. J. (2022). The impact on student achievement of an assessment for learning teacher professional development program. *Studies in Educational Evaluation*, 75, 101201. <https://doi.org/10.1016/j.stueduc.2022.101201>
- Feldman, I. (2025). The role of assessment in improving education and promoting educational equity. *Education Sciences*, 15(2), 224. <https://doi.org/10.3390/educsci15020224>
- Galle, S. A., Amos, G. N., & Tsaku, G. Y. (2024). Impact of table of specification on economics teachers' utilization of written curriculum implementation in secondary education of Nasarawa State, Nigeria. *Journal of Educational Research & Practice (JERP)*, 4(8). <https://berkeleypublications.com/bjerp/article/view/184>
- Garcia, M. L. B., Gaylo, D. N., & Vistro-Yu, C. P. (2024). Writing PISA-like mathematics items: The case of tertiary mathematics instructors from a state university in the Philippines. *European Journal of Educational Research*, 13(4), 1441–1453. <https://doi.org/10.12973/eu-jer.13.4.1443>
- Garcia, M. L. B., Santos, K. C., & Vistro-Yu, C. P. (2025). Comparing two psychometric approaches: The case of item analysis for a classroom test in mathematics. *International Journal of Education and Practice*, 13(1). <https://doi.org/10.18488/61.v13i1.4060>
- Gebremichael, M. W., Baraki, B., Mehari, M.-A., & Assalfew, B. (2025). Item analysis of multiple-choice questions from assessment of health sciences students, Tigray, Ethiopia. *BMC Medical Education*, 25(1), 441. <https://doi.org/10.1186/s12909-025-06904-6>
- Hacioglu, Y., & Gulhan, F. (2021). The effects of STEM education on the students' critical thinking skills and STEM perceptions. *Journal of Education in Science, Environment and Health (JESEH)*, 7(2), 139–155. <https://doi.org/10.21891/jeseh.771331>
- Hassan, M., Shaheen, R., & Ali, H. (2024). Evaluation of exam questions of English summative assessment based on cognitive skills. *Harf-O-Sukhan*, 8(2), 567–584. <https://harf-o-sukhan.com/index.php/Harf-o-sukhan/article/view/1375>
- Inayah, N., Muchyidin, A., & Toheri. (2024). Analysis of mathematics content knowledge of mathematics teachers based on Bloom's taxonomy revision. *Educational Insights*, 2(1), 79–87. <https://doi.org/10.58557/eduinsights.v2i1.36>
- Kazu, İ. Y., & Kurtoglu Yalçın, C. (2021). The effect of STEM education on academic performance: A meta-analysis study. *TOJET: The Turkish Online Journal of Educational Technology*, 20(4), 101–118. <https://www.researchgate.net>
- Khan, H. F., Qayyum, S., Beenish, H., Khan, R. A., Iltaf, S., & Faysal, L. R. (2025). Determining the alignment of assessment items with curriculum goals through document analysis by addressing identified item flaws. *BMC Medical Education*, 25(1), 200. <https://doi.org/10.1186/s12909-025-06736-4>
- Khan, R. A., & Aslam, K. (2024). A systematic approach to designing assessments: From rough estimate to purposeful weighting in table of specification and blueprinting. *Health Professions Educator Journal*, 7(2). <https://doi.org/10.53708/hpej.v7i2.3822>
- Khubaib Ahmed, Iqra Nazish, Tajammal Hussain, Yasir Abbas, Muhammad Haseeb, & Sheema khan. (2025). Identifying gaps in science curriculum student learning outcomes and Grade V students' achievement. *Indus Journal of Social Sciences*, 3(2), 243–258. <https://doi.org/10.59075/ijss.v3i2.1136>
- Kini, R. G., Marla, N. J., & Moras, C. B. (2025). Overcoming challenges in implementation of competency-based medical education in India: Leveraging postgraduates for multiple choice item design for undergraduate assessments in pathology. <https://doi.org/10.21203/rs.3.rs-5867554/v1>
- Kumar, R., & Lal, K. (2024). Teacher made test. *International Journal of Creative Research Thoughts*, 12(3), c227–c238. http://ijert.org/viewfull.php?&p_id=IJCRT2403279
- Lapinid, M. R. C., Mistades, V. M., Sagcal, R. R., Gustilo, L. E., Balagtas, M. U., Gonzales, R. D. L. C., Medula, C. T., & Palomar, B. C. (2024). Aligning Philippine K to 12 assessment policies against international benchmarks: Implications for quality reform. *Philippine Journal of Science*, 153(6B), 2375–2392. https://philjournalsci.dost.gov.ph/images/pdf/pjs_pdf/vol153_No6B_Dec2024/aligning_Philippine_K_to_12_assessment_policies_against Intl_benchmarks_.pdf
- Lemons, P. P., & Lemons, J. D. (2013). Questions for assessing higher-order cognitive skills: It's not just Bloom's. *CBE—Life Sciences Education*, 12(1), 47–58. <https://doi.org/10.1187/cbe.12-03-0024>

- Lestari, E. S., & Pratolo, B. W. (2025). Analysis of higher-order thinking skills (HOTS) in English final examination questions at junior high school. *Eternal: English Teaching Journal*, 16(1), 24–36. <https://doi.org/10.26877/Eternal.v16i1.892>
- Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessment, and instruction. *Review of Educational Research*, 79(4), 1332–1361. <https://doi.org/10.3102/0034654309341375>
- Murphy, D. H., Little, J. L., & Bjork, E. L. (2023). The value of using tests in education as tools for learning—not just for assessment. *Educational Psychology 1 Review*, 35(89). <https://doi.org/10.1007/s10648-023-09808-3>
- Orhani, S. (2024). Preparation of tests from the subject of mathematics according to Bloom's taxonomy. *International Journal of Research Publication and Reviews*, 5(2), Article 542. <https://doi.org/10.55248/gengpi.5.0224.0542>
- Raymundo, R. A. (2021). Evaluation of teacher-made examination in Grade 2 mathematics of Villaverde District. SSRN. <http://dx.doi.org/10.2139/ssrn.4950419>
- Rezkilaturahmi, I., Istiqlal, M., Putro, N. H. P. S., Istiyono, E., & Widiastuti. (2025). Evaluating English language test items developed by teachers: An item response theory approach. *Voices of English Language Education Society*, 9(1), 218–230. <https://doi.org/10.29408/veles.v9i1.27644>
- Şeker, Z. C., & Demirtojet, C. (2022). Investigation of the evaluation questions in 5th grade science textbooks according to revised Bloom's taxonomy. *The Turkish Online Journal of Educational Technology*, 21(4), 1–15. <https://files.eric.ed.gov/fulltext/ED638365.pdf>
- Shaikh, S., Daudpotta, S. M., & Imran, A. S. (2021). Bloom's learning outcomes' automatic classification using LSTM and pretrained word embeddings. *IEEE Access*, 9, 116162–116172. <https://doi.org/10.1109/ACCESS.2021.3106443>
- Stringer, J. K., Santen, S. A., Lee, E., Rawls, M., Bailey, J., Richards, A., Perera, R. A., & Biskobing, D. (2021). Examining Bloom's taxonomy in multiple-choice questions: Students' approach to questions. *Medical Science Educator*, 31, 1311–1317. <https://doi.org/10.1007/s40670-021-01305-y>
- Sukendra, I. (n.d.). Item analysis of English teacher-made test [Unpublished manuscript]. SMA Yaspita, Serpong Utara, Tangerang. <https://proceeding.unnes.ac.id/utnc/article/view/2627>
- Tiana, E. A., & Maruf, N. (2024). Aligning assessment and curriculum: A study of summative tests in educational settings. *JET (Journal of English Teaching) Adi Buana*, 9(1), 29–38. <https://jurnal.unipasby.ac.id>
- Zerman Kepceoğlu, G., & Pektaş, M. (2023). Analysis of sample science questions for secondary school exam in terms of revised Bloom taxonomy. *Kastamonu Education Journal*, 31(4), 567–575. <https://doi.org/10.24106/KEFDERGI-2023-0024>