

Assessing the Knowledge of Teachers in Objective Test Construction Procedure in the Teacher Education Programs: A Literature Review

Darryl D. Barrientos* Ham Jay M. Diano* Meshel D. Gamao* Kimberly Lepiten
College of Teacher Education, Cebu Roosevelt Memorial Colleges, San Vicente St.,
Bogo City, Cebu, Philippines 6010

* E-mail of the corresponding author: darryl.barrientos77@gmail.com

Abstract

This literature review aims to assess the knowledge of teachers in constructing objective tests in the teacher education programs where assessment of prospective teachers is seen as the most practical technique to improve and assess teacher candidates' abilities to make judgments that will help their students learn when presented with diverse scenarios in the classroom. In addition, in order to connect earlier studies on objective test construction, it is essential to highlight literature on Multiple-Choice test and the Revised Bloom's Taxonomy which is seen as a guide in constructing variety of valid and reliable objective tests. Doing so will enable to make connections on the procedure and rules of constructing objective tests for prospective teachers.

Keywords: objective test construction procedure, multiple-choice, revised bloom's taxonomy, assessment

DOI: 10.7176/JEP/14-24-06

Publication date: August 31st 2023

1. Introduction

One of the most popular educational aims is to help students develop higher-order thinking skills (Mainali, 2012). To assess students' higher-order thinking, it is critical to choose the correct assessment technique (Kaipa, 2020). Indeed, it has been proven that students exposed to tests that require higher-order thinking are more likely to adopt meaningful, holistic approaches to their studies rather than relying on rote learning techniques (Jensen, McDaniel, Woodard, & Kummer, 2014). Furthermore, such assessments allow teachers to provide more precise and specific feedback, which can help to stimulate and steer future learning (Scully, 2017). A test is commonly used in schools as an assessment tool to acquire information about students' learning (Quansah, Amoako, & Ankomah, 2019). To understand more deeply the process of constructing a valid test to assess student's higher-order thinking skills, such constructs will be explained: (1) Revised Bloom's Taxonomy, (2) Teacher's role toward proper assessment tool, (3) Objective test to assess student thinking skills and learning (4) Multiple-choice objective test, (5) Challenges in the construction of accurate multiple-choice tests, (6) Developing test specifications, (7) Selecting appropriate item types, (8) Preparing relevant test items, and (9) Assembling the test.

Teachers create and give exams at different levels of education worldwide to measure the amount of learning and abilities that students have acquired (Quaigrain & Arhin, 2017). Higher education has a strong interest in assessing higher-order skills. Universities and third-level institutions are under increasing pressure to close the gap between what students learn and what employers value (Scully, 2017). Some criticize the current state of assessment in higher education, claiming that it has little impact on educational quality and that accrediting organizations demand schools devote time and resources to gathering information on student learning even if it does not increase academic quality (Gilbert, 2015). However, well-crafted test items should be used to assess what students know or have learned in a particular subject area (Quansah, Amoako, & Ankomah, 2019). The assessment gives information on which educational decisions are made, such as the success of learning programs or students who have achieved specific levels of competence and knowledge (Agu, Onyekuba, & Anyichie, 2013). The assessment procedures used have a significant impact on the learning quality (Fernandes, Flores, & Lima, 2012). They can affect how students think about learning (Pereira, Flores, & Niklasson, 2016). Assessment approaches based on a learner-centered assessment increase students' active participation, create feedback, allow students and faculty to collaborate, and enable teachers to see how learning occurs (Webber, 2012). These practices help learners prepare for careers by encouraging problem-solving and skill development in real-world situations (Pereira, Flores, & Niklasson, 2016). The necessity to assess students' growth and the requirement to give approved public qualifications make student assessment a top priority for educators (Pittaway, 2012).

1.1 Revised Bloom's Taxonomy

Students lack the ability to adapt what they've learned in one context to a new one. Our educational institutions teach students to be decent citizens and industrial workers. Students were required to sit, listen, and follow instructions to the letter. In some aspects, this paradigm benefited graduates from schools and universities since

they learned to follow directions in ways that would help their future jobs (Mainali, 2012). Moreover, pedagogy in the classroom may profoundly influence the style of thought in which students participate in schools and colleges. We often become so focused on 'what' kids must learn that we neglect 'how' best to guarantee they learn. It considers how students' minds react to the learning environment established by their teachers and peers (Mainali, 2012).

Bloom's taxonomy was recently amended to increase its instructional value and accuracy (Anderson & Krathwohl, 2001). While all taxonomies listed above have been developed and used for many years, a new version of the cognitive taxonomy, known as Bloom's Taxonomy, emerged in the twenty-first century (Wilson, 2016).

Before it was instituted in an updated version (Anderson & Krathwohl, 2001), there have been efforts to break down and define the many domains of human learning - cognitive (knowing, or brain), affective (feelings, or heart), and psychomotor (doing, or kinesthetic, tactile, haptic, or hand/body). As a result of their work, each area now has its taxonomies. The taxonomies stated above deal with various elements of human learning and are organized hierarchically, starting with the essential functions and progressing to the more complicated ones (Wilson, 2016).

Anderson changed Bloom's categories from nouns to verbs, altering the original wording (Darwazeh & Branch, 2015). Anderson modified the knowledge, comprehension, and synthesis categories to remember, understand, and create. Anderson also rearranged the synthesis order, putting it at the top of the triangle under the term Create. Thus, the updated Bloom's taxonomy (Anderson & Krathwohl, 2001) became: Remember, Understand, Apply, Analyze, Evaluate, and Create.

Academics assist students in achieving the skills and understanding needed by matching instructional methods, assessment, and the classroom environment in the form of constructive alignment (Sagala & Andriani, 2019). Nowadays, education is expected to take students beyond memorization and recall of facts. Because the amount of information and facts available is rapidly rising, students will be unable to compete in this environment if they cannot comprehend, analyze, apply, evaluate, and create (Crossland, 2015). Based on Bloom's Taxonomy (1965), these different levels of cognitive skills are divided into two groups (Qasrawi & BeniAbdelrahman, 2020): lower-order thinking skills (LOTS) and higher-order thinking skills (HOTS).

We cannot develop an examination employing several LOTS to measure the gained skills of final-year students academically (simple recall of information). Similarly, because they are still absorbing new material, first-year students cannot be expected to answer numerous HOTS (assessment of complex issues). As a result, proper attention must be paid to test papers to preserve the appropriate balance of lower and higher-order thinking skills (Sagala & Andriani, 2019).

Students address difficulties at the application level by immediately applying the information or knowledge they have learned. At the analysis stage, students must be able to break down a whole into pieces and figure out how the parts are related to form a whole. Students must be able to make judgments based on particular criteria and standards at the assessment level. Because children are claimed to be able to build new goods by remodeling certain pieces or parts to a shape or structure that the instructor has never discussed before, the degree of synthesis is also known as creative behavior. Synthesis is referred to as creating in the previous description (Pratama & Retnawati, 2018).

Education reform entails being updated to the abilities that learners will require to meet the demands of the twenty-first century. Innovation, life and career skills and technological skills are among the expectations. Notably, such expectations necessitate learners' ability to communicate, collaborate, think critically, and be creative, among other things (Qasrawi & BeniAbdelrahman, 2020). Thus, abstract and concrete skills are the two basic 21st-century skills (Rentawati, Djidu, Apino, & Anazifa, 2018). It's worth noting that higher-order thinking abilities go under the abstract category, while communication and teamwork fall under the concrete category. Furthermore, the development of HOTS is linked to the development of creative and critical thinking skills (Qasrawi & BeniAbdelrahman, 2020). The dedication to HOTS coincided with the advancement of information and technology, where learners require various skills to deal with vast amounts of data, such as analysis, evaluation, and creation (Halili, 2015). Some scholars feel that HOTS are also crucial in developing lifelong learning, which helps learners effectively adapt to the demands of the twenty-first century (Rentawati, Djidu, Apino, & Anazifa, 2018).

1.2 Teachers' Role Toward Valid Assessment Tool

Given the value of test scores, the importance of teachers creating proper examinations for their students is unquestionable (Inko-Tariah & Okon, 2019). According to researchers, a teacher's competency significantly impacts the quality of reviews created (Darling-Hammond, 2012). However, those who see teaching and learning as the transmission of knowledge from teacher to student are more likely to see assessment to test students' ability to replicate information (Fletcher, Meyer, Anderson, Johnston, & Rees, 2012). Those who believe that teaching and learning should facilitate critical thinking and knowledge transfer, on the other hand, consider

evaluation as an essential element of the learning process. As a result, assessment activities, especially in traditional higher education, are still conducted as add-ons to the curriculum, designed for program evaluation rather than student mastery (Gyll & Ragland, 2018).

Teachers are primarily responsible for leading and carrying out most of the key responsibilities and activities in educational institutions. Hence the quality of the education system and the profession of educators are heavily reliant on them (Swarnalatha, 2016). Teachers play an essential role in assisting and encouraging students to learn in the classroom (Ankomah, 2020). Teachers should be dedicated to their job and demonstrate that they are worthy of the public's trust and confidence by offering high-quality education to all students. Teachers dedicated to their careers have higher efficacy, job satisfaction, and competence levels. The true teacher strives for higher performance and stays up with the latest abilities used in the classroom to teach learning content (Swarnalatha, 2016).

Test construction necessitates applying abilities that allow a teacher to develop a test with precision, including acceptable language use, objective communication, item validation, and appropriate grading scales. Still, they must establish broad test construction skills to ensure that items are structured so that they elicit obvious and detectable variations among learners exclusively on the constructs being assessed (Ankomah, 2020). Good assessments aren't one-time events; they're part of a long-term, systematic, and coordinated endeavor to better understand and enhance teaching and learning (Gyll & Ragland, 2018) and teachers' lack of test-construction skills may lead to erroneous assessments of students' achievements (Agu, Onyekuba, & Anyichie, 2013). If learning and instructional objectives are to be met effectively, every instructor must be proficient in test construction (Quansah et al., 2019).

Unfortunately, the role of instructors in test construction has been identified as a significant cause of anxiety, especially with teachers who have only a few years of experience in the classroom, and the lack of test building abilities of these teachers is partly to blame for this worry (Quansah et al., 2019). Scholars, likewise, have argued that teachers' use of tests is not encouraging (Hamafyelto, Hamman-Tukur, & Hamafyelto, 2015). The implication is that teachers may accept incorrect information about student learning (Quansah et al., 2019).

1.3 Objective Test to Assess Student Thinking Skills and Learning

In recent years, educators have become more aware of the potential importance of formative assessment in education (Scully, 2017) for giving instructors feedback on the strengths and weaknesses of their students' learning and offering student growth evaluations. Formative assessment is a term that refers to assessments that are used to improve the teaching-learning process (Hortigüela, Palacios, & López, 2018). This aim should be met by using exams that can assess students' HOTS (Eka Mahendra, Parmithi, Hermawan, Putu Juwana, & Gunartha, 2020).

An objective exam is a tool for assessing a student's progress toward a goal (Rahmawati, Suwandi, Saddhono, & Setiawan, 2019). In an objective test, the testee must choose the correct answer from various options (Igbojinwaekwu, 2015). Objective tests have the following advantages: quick, accurate, and repeatable. However, to conduct an objective test, criteria that demand subjective judgment and value must be met (Singham, Birwal, & Yadav, 2015). Test items are independent in an objective test, and the ability to correctly answer one question is unrelated to the capacity to answer another (Fox, 2012) correctly.

For a competency-based assessment program, objective assessments are an effective alternative, and this sort of assessment typically consists of multiple-choice, matching, and short-answer items that can be scored and delivered to students rapidly, which is advantageous for students who like to work at their own pace (Gyll & Ragland, 2018). In practice, teacher-made assessments are still infrequently used to examine students' HOTS as formative evaluation (Mahendra, Jayantika, & Sulistyani, 2019). Assessing students' high-order thinking capabilities can help them improve their cognitive abilities (Mahendra, Jayantika, & Sulistyani, 2019).

1.4 Multiple-Choice Objective Test

Multiple-choice is regarded as the most adaptable and useful of the objective exam types (Alade & Igbiosa, 2014). However, it is the most challenging exam format (Rahmawati, Suwandi, Saddhono, & Setiawan, 2019). It's possible that the problem stems from the presence of diversions among the responses (Mahjabeen et al., 2017). Thus, it requires exact knowledge (Rahmawati et al., 2019) and comprehension to answer the questions. MCQs need respondents to choose the proper answer from various options, making it easier to evaluate and deliver quick feedback in a large classroom (Mullen & Schultz, 2012). When students prepare for an exam that consists of MCQs, it is easier for them to recollect the essential substance of the lecture rather than the specifics (Kaipa, 2020). Compared to other assessment methods, MCQs tend to be more objective (Kaipa, 2020).

Testing is frequently thought of as an assessment tool, but it is also a learning exercise (Pan & Rickard, 2018). Retrieving information in response to a test question enhances memory, resulting in higher long-term retention; it can also change the representation of the material in memory, resulting in deeper comprehension. Importantly, similar favorable impacts of multiple-choice testing can also be shown in real-world educational

situations (Butler, 2018). Multiple-choice testing, for example, has been conducted to improve retention and transfer on the subsequent unit and final exams in middle school (McDaniel, Thomas, Agarwal, McDermott, & Roediger, 2013), high school (McDermott, Agarwal, D'Antonio, Henry L. Roediger, & McDaniel, 2014), and college courses (McDaniel, Wildman, & Anderson, 2012). In addition, multiple-choice testing can help students learn non-tested, conceptually related information (Bjork, Little, & Storm, 2014) and reclaim previously acquired knowledge that has been inaccessible (Butler, 2018).

Planning, item writing, item analysis, item composition, test theory, reliability, printing, and manual preparation are all used to design a test (Osadebe, 2015). These processes determine the test's content area, format, and table specifications (Rivai, Ridwan, Supriyati, & Rahmawati, 2019). The principles of test development, administration, analysis, and reporting must all be strictly followed (Quaigrain & Arhin, 2017).

1.5 Challenges in the Construction of Multiple-Choice Objective Test

Adherence to standard procedures for test construction is required for high-quality classroom-based assessment. Every classroom instructor is expected to have and use the necessary abilities to create high-quality items for class evaluations (Agu, Onyekuba, & Anyichie, 2013). Students have issues with teacher-made tests characterized by the invalidity, overtesting, insufficient administration time, test items that do not cover course topics, etc. This indicates that the test was not content valid (Alade & Igbinsosa, 2014).

Fairness, ease, anxiety, and performance have all been discussed in Multiple-Choice (MC) exams in some articles (Núñez-Peña & Bono, 2021). Students voiced worries about fairness not regarding whether they received a printed copy of the exam but in relation to their overall course achievement (Emeka & Zilles, 2020).

There are several challenges to developing and using MCQs, including the tendency to write poor MCQs with ambiguous prompts, poor distractors, multiple answers when only one correct answer is required, controversial answers, give-away keys, and a higher probability of testees guessing correctly, to name a few (Odukoya, Adekeye, Igbinsoba, & Afolabi, 2018).

1.6 Developing Test Specifications

A multiple-choice item is made up of stem, which includes information such as context, content, and/or the question that the student must answer (Gierl, Bulut, Guo, & Zhang, 2017); the options, which are just as vital as the stem and consists of the appropriate option and incorrect options or distractors; and any additional information which includes any other content either in the stem or options. The branch and suitable options are the most crucial parts of the multiple-choice item for most test developers and users. Meanwhile, distractors add to the context needed to solve a multiple-choice item, potentially affecting item quality and learning results (Gierl et al., 2017). The effects of partial knowledge in response performance, which interacts with the plausibility of each distractor to affect the psychometric properties of the correct and incorrect options, foster this complex relationship (Haladyna & Rodriguez, 2013). The benefits of the testing effect, on the other hand, are dependent on the quality of the distractors. When the information relating to the correct option and the distractors are evaluated on subsequent exams, competitive multiple-choice items (i.e., items where the distractors are plausible and share essential information with the right option) elicit advantageous retrieval processes (Little & Bjork, 2015).

Meanwhile, to help teachers align assessment to the objectives and instructions, a table of specifications, often known as a test blueprint, can be constructed (Alade & Igbinsosa, 2014). A table of specifications is an activity that lists the knowledge and cognitive tasks on which examinees will be evaluated. It is as well defined as possible scope, which establishes the test's focus and links other objectives to the content to provide a balanced set of test items (Alade & Igbinsosa, 2014).

There are six primary aspects to consider when creating a table of specifications for a comprehensive test (Alade & Igbinsosa, 2014) which include: Balance among the objects that have been chosen for the test. Balance among levels of learning. Format of the test. A total number of items. A number of items for each objective and level of learning. Selection of skills from each objective framework.

This table of specifications for classroom application is to assist classroom instructors in developing assessments that are adequately linked to the subject matter covered and the cognitive process employed during instruction. However, for this method to be, teachers must make it their own and conduct a practical assessment (Alade & Igbinsosa, 2014).

1.7 Selecting Appropriate Item Types

It has been proven that students who are exposed to tests that require higher-order thinking are more likely to adopt meaningful, holistic approaches to their studies rather than relying on rote learning strategies (Jensen, McDaniel, Woodard, & Kummer, 2014). Furthermore, such tests enable teachers to offer more precise and specific feedback, which can help to stimulate and steer future learning (Scully, 2017). Because they appear to lack test creation abilities, some teachers create substandard tests, while others continue to employ replicas of

test objects (Ankomah, 2020). Likewise, teachers make exam items that solely measure lower-order cognitive skills (Hamafyelto, Hamman-Tukur, & Hamafyelto, 2015).

The validity of derived scores is required for an assessment instrument, regardless of the degree of education of the examinees or the domain or topic being assessed (Ali, Carr, & Ruit, 2016). An assessment of a student's achievement can also serve as feedback on a teacher's performance (Rahmawati, Suwandi, Saddhono, & Setiawan, 2019). Students are notoriously weak at multiple-choice tests and strong on other test formats, which implies that teachers should assess their teaching approaches to increase students' capacity to recognize the distractor in a multiple-choice test format (Rahmawati, Suwandi, Saddhono, & Setiawan, 2019).

1.8 Preparing Relevant Test Items

In a study conducted, it was clear that the test developer, who is also a subject instructor, did not sample enough to cover all of the content areas stated in the relevant term's scheme of work (Quansah, Amoako, & Ankomah, 2019). According to an analysis of the papers, the substance of these articles focused on a handful of the subjects taught. The items on the instruments (tests) were insufficiently representative of the content prepared.

The significant material, skills, and learning outcomes stated in the school's or district's curricular framework and content standards are unlikely to be reflected in an assessment task that lacks content validity (Quansah, Amoako, & Ankomah, 2019). If an MCQ is to be used to measure higher-order cognitive skills, there must be a mechanism that provides proper training and feedback to the item creators (Khan, Danish, Awan, & Anwar, 2013). The findings show that the standard of MCQs for assessment papers can be improved by repetition and practice. To ensure a higher quality of MCQ, it is recommended that the items be examined at an inter-departmental level before being submitted to the finalizing committee. This will result in better-authored materials while also saving time.

In addition, technical item defects are common during MCQ development, and identifying these flaws leads to increased quality of the single best MCQs. To correct these flaws, faculty should be trained in item writing abilities (Khan, Danish, Awan, & Anwar, 2013).

1.9 Assembling the Test

During the early item development process, item unidimensionality used to describe items or test scores frequently plays a crucial role, either directly or implicitly (Ziegler & Hagemann, 2015). Scoring generates raw scores, which indicate the number of correct answers or points given to students (Pter & Line, 2015). Multiple-choice tests have traditionally been graded using the number right (NR) scoring technique. Correct answers receive a positive score, whereas erroneous responses and absent or missing answers receive a zero. The test score is the total of the valid response scores. (Lesage, Valcke, & Sabbe, 2013).

Several large-scale educational testing systems compose various test formats with minimum or no overlap to safeguard the exam and maintain its validity. At the same time, test takers should be unconcerned about the administered exam form (Debeer, Ali, & Rijn, 2017). As a result, test forms that will be utilized parallel should have the same statistical and content-related features (Chen, Chang, & Wu, 2012). Automatic test assembly (ATA), or the automatic selection of things from an item pool to build one or more new test forms, is critical in developing parallel test forms (Debeer, Ali, & Rijn, 2017).

Moreover, the test should be carried out in a safe and secure atmosphere (Pter & Line, 2015) to ensure that students are comfortable for them to focus on the analysis and comprehension of test items. The setting should be set up with appropriate lighting, a pleasant temperature, and enough workspace. Test items and instructions should also be supplied to students (Pter & Line, 2015).

2. Synthesis

Assessment is considered one of the most important teaching and learning processes. This allows teachers to evaluate the students' learning of the content delivered in teaching and the achievement of the objectives. Likewise, this allows educational administrators and the institution to gauge student achievement and success. Teacher Education programs in the Philippines are board courses requiring graduates to take the Licensure Examination for Teachers (LET), stipulated in the Republic Act 7836 or the Philippine Teachers Professionalization Act of 1994. This is one of the highlighted standards to qualify as a professional teacher. This standard objective examination is conducted using Multiple-choice items. Thus, teacher education students must be provided with assessments that would prepare them for the qualifications stated in the Act. This brings the responsibility of the teachers in the Teacher Education programs to utilize their knowledge in objective test construction procedures considering that this is the best way to prepare prospective teachers to be qualified to be in the field of the academe.

Multiple-choice questions may appear to be easy considering their structure wherein the answer, together with some other options called distractors, are presented along with the questions, but this is one of the most complex types of tests to construct. Teachers must be well-equipped with the skills to build any helpful

assessment tool. The knowledge and utilization of assessment instruments are keys to competence in assessment. According to studies, instructors' ability, particularly in review, impacts the test's quality (Darling-Hammond, 2012). To develop accurate assessment instruments, all stakeholders in the educational system need excellent training in test creation skills (Ankomah, 2020). Graduation from a higher education institution does not imply competency in assessment skills. Being a classroom teacher and administering tests does not provide the necessary skills for developing valid and credible assessment tools.

In this case, the teachers' knowledge of test construction procedure, one of the most critical aspects of teaching and learning, can never be neglected. Schools and institutions should ensure that teachers exhibit competence in test construction to provide quality education, considering that assessments enable teachers and schools to evaluate student learning and achievement.

References

- Agu, N. N., Onyekuba, C., & Anyichie, A. C. (2013). Measuring teachers competencies in constructing classroom-based tests in Nigerian secondary schools: Need for a test construction skill inventory. *Educational Research and Reviews*, 8(8), 431-439.
- Alade, O., & Igbinsola, V. O. (2014). Table of specification and its relevance in educational development assessment. *European Journal of Educational and Development Psychology*, 2(1), 1-17.
- Ali, S. H., Carr, P. A., & Ruit, K. G. (2016). Validity and Reliability of Scores Obtained on Multiple-Choice Questions: Why Functioning Distractors Matter. *Journal of the Scholarship of Teaching and Learning*, 16(1), 1-14.
- Anderson, L. W., & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Allyn & Bacon.
- Ankomah, F. (2020). Predictors of adherence to test construction principles: The case of senior high school teachers in Sekondi-Takoradi Metropolis (Doctoral dissertation, University of Cape Coast).
- Bjork, E. L., Little, J. L., & Storm, B. C. (2014). Multiple-choice testing as a desirable difficulty in the classroom. *Journal of Applied Research in Memory and Cognition*, 3(3), 165-170.
- Butler, A. (2018). Multiple-choice testing in education: Are the best practices for assessment also good for learning? *Journal of Applied Research in Memory and Cognition*, 7(3), 323-331.
- Chen, P.-H., Chang, H.-H., & Wu, H. (2012). Item selection for the development of parallel forms from an IRT-based seed test using a sampling and classification approach. *Educational and Psychological Measurement*, 72, 933-953.
- Crossland, J. (2015). Thinking Skills and Bloom's Taxonomy. *Primary Science*, 32-34.
- Darling-Hammond, L. (2012). *Powerful teacher education: Lessons from exemplary programs*. John Wiley & Sons.
- Darwazeh, A. N., & Branch, R. M. (2015). A revision to the revised Bloom's taxonomy. *Annual Proceedings-Indianapolis*, 220-226.
- Debeer, D., Ali, U. S., & Rijn, P. W. (2017). Evaluating statistical targets for assembling parallel mixed-format test forms. *Journal of Educational Measurement*, 54(2), 218-242.
- Eka Mahendra, I. W., Parmithi, N. N., Hermawan, E., Putu Juwana, D., & Gunartha, I. W. (2020). Teachers' Formative Assessment: Accessing Students' High Order Thinking Skills (HOTS). *International Journal of Innovation, Creativity and Change*, 12(12), 180-202.
- Emeka, C., & Zilles, C. (2020). Student Perceptions of Fairness and Security in a Versioned Programming Exam. In *Proceedings of the 2020 ACM Conference on International Computing Education Research*, 25-35.
- Fernandes, S., Flores, M. A., & Lima, R. M. (2012). Students' views of assessment in project-led engineering education: findings from a case study in Portugal. *Assessment & Evaluation in Higher Education*, 37(2), 163-178.
- Fletcher, R. B., Meyer, L. H., Anderson, H., Johnston, P., & Rees, M. (2012). Faculty and students conceptions of assessment in higher education. *Higher education*, 64(1), 119-133.
- Fox, J. (2012). Language assessment methods. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics*. Oxford, UK: Blackwell Publishing Ltd. doi:<http://doi.org/10.1002/9781405198431.wbeal0606>
- Gierl, M. J., Bulut, O., Guo, Q., & Zhang, X. (2017). Developing, analyzing, and using distractors for multiple-choice tests in education: a comprehensive review. *Review of Educational Research*, 87(6), 1082-1116.
- Gilbert, E. (2015). Does assessment make colleges better? If so, we haven't seen the evidence. *The Chronicle of Higher Education*, 62, 50-51.
- Gyll, S., & Ragland, S. (2018). Improving the validity of objective assessment in higher education: Steps for building a best-in-class competency-based assessment program. Wiley, 1-8. Retrieved from Wiley.
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. New York: Routledge.
- Hamafyelto, R. S., Hamman-Tukur, A., & Hamafyelto, S. S. (2015). Assessing teacher competence in test construction and content validity of teacher made examination questions in commerce in Borno State,

- Nigeria. *Journal of Education*, 5(5), 123-128.
- Hortigüela, A. D., Palacios, P. A., & López, P. V. (2018). The impact of formative and shared or co-assessment on the interview requisition of transversal competences in shared or co-assessment on the interview requisition of transversal competences in. *Assessment & Evaluation in Higher Education*, 1-13. doi:10.1016/j.compedu.2010.12.002
- Igbojinwaekwu, P. C. (2015). Effectiveness of guided multiple choice objective questions test on students' academic achievement in senior school Mathematics by school location. *Journal of Education and Practice*, 6(11), 37-48.
- Inko-Tariah, D. C., & Okon, E. J. (2019). Knowledge of test construction procedures among lecturers in Ignatius Ajuru University of Education, Port Harcourt, Nigeria. *Academic Research International*, 10(1), 130-138.
- Jensen, J. L., McDaniel, M. A., Woodard, S. M., & Kummer, T. A. (2014). Teaching to the test... or testing to teach: Exams requiring higher order thinking skills encourage greater conceptual understanding. *Educational Psychology Review*, 26(2), 307-329.
- Kaipa, R. M. (2020). Multiple choice questions and essay questions in curriculum. *Journal of Applied Research in Higher Education*.
- Khan, H. F., Danish, K. F., Awan, A. S., & Anwar, M. (2013). Identification of technical item flaws leads to improvement of the quality of single best multiple choice questions. *Pakistan journal of medical sciences*, 29(3), 715.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice*, 41(4), 212-218.
- Lesage, E., Valcke, M., & Sabbe, E. (2013). Scoring methods for multiple choice assessment in higher education—Is it still a matter of number right scoring or negative marking? *Studies in Educational Evaluation*, 39(3), 188-193.
- Little, J. L., & Bjork, E. L. (2015). Optimizing multiple-choice tests as tools for learning. *Memory & Cognition*, 43(1), 14-26.
- Mahendra, W. E., Jayantika, I. G., & Sulistyani, W. R. (2019). Developing HOTS through performance assessment. *International Journal of Scientific & Technology Research*, 8(11), 3004-3007.
- Mahjabeen, W., Alam, S., Hassan, U., Zafar, T., Butt, R., Konain, S., & Rizvi, M. (2017). Difficulty index, discrimination index and distractor efficiency in multiple choice questions. *Annals of PIMS-Shaheed Zulfiqar Ali Bhutto Medical University*, 13(4), 310-315.
- Mainali, B. P. (2012). Higher order thinking in education. *Academic Voices: A Multidisciplinary Journal*, 2(1), 5-10.
- McDaniel, M. A., Thomas, R. C., Agarwal, P. K., McDermott, K. B., & Roediger, H. L. (2013). Quizzing in middle-school science: Successful transfer performance on classroom exams. *Applied Cognitive Psychology*, 27(3), 360-372.
- McDaniel, M. A., Wildman, K. M., & Anderson, J. L. (2012). Using quizzes to enhance summative-assessment performance in a web-based class: An experimental study. *Journal of Applied Research in Memory and Cognition*, 1(1), 18-26.
- McDermott, K. B., Agarwal, P. K., D'Antonio, L., Henry L. Roediger, I., & McDaniel, M. A. (2014). Classroom-based programs of retrieval practice reduce middle school and high school students' test anxiety. *Journal of Applied Research in Memory and Cognition*, 3(3), 131-139.
- Mullen, K., & Schultz, M. (2012). Short answer versus multiple choice examination questions for first year chemistry. *International Journal of Innovation in Science and Mathematics Education*, 20(3).
- Núñez-Peña, M. I., & Bono, R. (2021). Math anxiety and perfectionistic concerns in multiple-choice assessment. *Assessment & Evaluation in Higher Education*, 46(6), 865-878.
- Odukoya, J. A., Adekeye, O., Igbinoba, A. O., & Afolabi, A. (2018). Item analysis of university-wide multiple choice objective examinations: the experience of a Nigerian private university. *Quality & quantity*, 52(3), 983-997.
- Osadebe, P. U. (2015). Construction of Valid and Reliable Test for Assessment of Students. *Journal of Education and Practice*, 6(1), 51-56.
- Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning: Meta-analytic review and synthesis. *Psychological Bulletin*, 144(7), 710.
- Pereira, D., Flores, M. A., & Niklasson, L. (2016). Assessment revisited: a review of research in Assessment and Evaluation in Higher Education. *Assessment & Evaluation in Higher Education*, 41(7), 1008-1032.
- Pittaway, S. M. (2012). Student and staff engagement: developing an engagement framework in a faculty of education. *Australian Journal of Teacher Education*, 37(4), 37-45.
- Pratama, G. S., & Retnawati, H. (2018). Urgency of higher order thinking skills (HOTS) content analysis in mathematics textbook. In *Journal of Physics: Conference Series*, 1097(1).
- Pter, C., & Line, O. (2015). Assembling, Administering, Scoring, and Reporting a Test. *Textbook of Nursing Education-E-Book*.

- 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*, 7(3), 744-758.
- Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, 4(1).
- Quansah, F., Amoako, I., & Ankomah, F. (2019). Teachers' test construction skills in Senior High Schools in Ghana: Document analysis. *International Journal of Assessment Tools in Education*, 6(1), 1-8.
- Rahmawati, L. E., Suwandi, S., Saddhono, K., & Setiawan, B. (2019). Construction of test instrument to assess foreign student's competence. *International Journal of Instruction*, 12(4), 35-48.
- Rentawati, H., Djidu, H., Apino, K., & Anazifa, R. (2018). Teachers' knowledge about higher-order thinking skills and its learning strategy. *Problems of Education in 21st Century*, 76(2).
- Rivai, E., Ridwan, A., Supriyati, Y., & Rahmawati, Y. (2019). Influence of test construction knowledge, teaching material and attitude on sociological subject to quality of objective test in public and private vocational schools. *International Journal of Instruction*, 12(3), 497-512.
- Sagala, P. N., & Andriani, A. (2019). Development of higher-order thinking skills (HOTS) questions of probability theory subject based on bloom's taxonomy. In *Journal of Physics: Conference Series*, 1118(1).
- Scully, D. (2017). Constructing multiple-choice items to measure higher-order thinking. *Practical Assessment, Research, and Evaluation*, 22(1), 4.
- Singham, P., Birwal, P., & Yadav, B. K. (2015). Importance of objective and subjective measurement of food quality and their inter-relationship. *J Food Process Technol*, 6(9), 1-7.
- Swarnalatha, S. S. (2016). Work commitment of secondary school teachers. *International Journal of Indian Psychology*, 3(4), 84-89.
- Webber, K. L. (2012). The use of learner-centered assessment in US colleges and universities. *Research in Higher Education*, 53(2), 201-228.
- Wilson, L. O. (2016). Anderson and Krathwohl–Bloom's taxonomy revised. *Understanding the New Version of Bloom's Taxonomy*.
- Ziegler, M., & Hagemann, D. (2015). Testing the unidimensionality of items. *European Journal of Psychological Assessment*, 31(4), 231–237.