

USE OF TEST BLUEPRINT IN IMPROVING TEACHERS' TEST CONSTRUCTION SKILLS FOR QUALITY ASSESSMENT

Obilor, Esezi Isaac & Omeke, Kasitem Sheila

Submitted: 28 December, 2023 Revised: 5 June, 2024 Accepted: 7 June, 2024

Abstract

This study examined the use of test blueprint in improving teachers' test construction skills for quality assessment. Test blueprint is an important and effective tool for quality test construction. Test blueprints provide a systematic approach to test development, ensuring that assessments are well-constructed, reliable, and valid measures of students' learning outcomes. By using a test blueprint, teachers can ensure that their assessments are closely aligned with learning objectives, provide appropriate test item selection, and provide a balanced coverage of content. The use of test blueprints can help teachers create assessments that provide accurate measures of students' learning outcomes and useful feedback to students. By providing a systematic approach to assessment development, teachers can improve the reliability, validity, and fairness of their assessments, which ultimately results in more accurate measurements of students' learning experiences, leading to quality assessment of students' learning outcomes. The study concluded that Test blueprints are crucial for enhancing teachers' test construction skills, ensuring tests are well-constructed, align with learning objectives, and provide accurate measures and feedback, thereby improving the reliability and validity of assessments. The study suggested among others that teachers should have a clear understanding of the test blueprint, which outlines the content, objectives, and skills that a test should assess; teachers should use test blueprint as a guide for developing test items that accurately assess the intended knowledge and skills of learners; teachers should continuously review their test blueprints to ensure that they are still relevant and valid.

Keywords: Assessment, blueprint, construction, skills, test.

Introduction

Test is a procedure or a series of questions or tasks designed to assess a person's knowledge, skills, abilities, or performance in a particular domain or subject. According to the American Psychological Association (APA), a test is an objective and standardized measure of a sample of behaviour (APA, 2020). Tests can be administered in a variety of formats, including written or oral, online or offline, and can be used for a range of purposes such as psychological assessment, educational assessment, diagnostic assessment, employment assessment, and more. Psychological assessment refers to the process of measuring and evaluating various aspects of an individual's psychological functioning. These assessments can be conducted for a variety of reasons, including diagnosing mental health conditions, determining eligibility for certain services or programmes, and informing treatment planning. There are many different types of psychological assessments that can be used, ranging from structured interviews and self-report questionnaires to standardized tests and behavioural observations (Herman & Golan, 1991). Some of the most commonly used standardised tests include the Minnesota Multiphasic Personality Inventory (MMPI), the Rorschach Inkblot Test, and the Wechsler Adult Intelligence Scale (WAIS) (Gutiérrez *et al.*, 2020).

One of the primary benefits of psychological assessment is that it can provide objective data on an individual's psychological functioning. This information can be used to inform treatment decisions, monitor progress over time, and help individuals better understand their own thoughts, feelings, and behaviours. However, it

is important to note that psychological assessments are not infallible and should always be interpreted within the context of the individual's unique experiences and circumstances (Quansah *et al.*, 2019). Educational assessment is the process of gathering and analysing information about students' learning and academic performance. It plays a critical role in educational systems worldwide, helping educators and policymakers to make informed decisions about curriculum, teaching methods, and resource allocation (Stiggins, 2020). With new research and advancements in technology, educational assessment practices have evolved significantly in recent years. One form of educational assessment is formative assessment.

Formative assessment, a continual process, harnesses evidence of student learning to refine instruction and tailor it to their needs. It embraces diverse methods like classroom observations, informal questioning, peer, and self-assessment, furnishing immediate feedback. This approach, as highlighted by Wiliam (2018), significantly enhances academic progress. Classroom observations, integral to this process, aid in refining test construction practices. They unveil insights into test validity and reliability by scrutinizing student interactions with test items. Additionally, observations help unveil biases, ensuring fairness across diverse student groups (Krouska *et al.*, 2018). Moreover, they offer invaluable feedback, refining the clarity and comprehensibility of test items (Popham, 2018).

Black and Wiliam (2018) agreed that educational assessment is pivotal in enhancing students' learning outcomes by identifying strengths and areas for improvement. Modern assessment practices emphasize relevance and engagement, leveraging technology for personalized learning experiences. However, when traditional assessments fail to address specific learning challenges, diagnostic assessments are essential. They delve into the root causes of learning difficulties, allowing educators to design targeted interventions. For instance, if a student consistently struggles with reading comprehension, a diagnostic assessment can uncover whether the issue stems from vocabulary deficits, decoding skills, or

background knowledge. This precise identification enables tailored support, ensuring that each student receives the necessary resources to overcome obstacles and succeed academically.

Diagnostic assessment is a form of evaluation used by educators to determine learner's strengths and weaknesses in a particular area or skill. The goal of diagnostic assessment is to identify specific areas in which the learner may need further support or intervention. Diagnostic assessments are typically administered before a formal intervention or treatment plan is developed, and they may be used to inform the development of individualised education plans (IEPs) or treatment plans. Diagnostic assessment can be used in a variety of settings, including schools, healthcare facilities, and workplaces. In schools, diagnostic assessments are often used to identify students who may need additional support or intervention, such as those who are struggling with reading or math. It may also be used in healthcare settings to evaluate patients who are experiencing health, cognitive or behavioural issues. Diagnostic assessment is an important tool for ensuring that individuals receive the support and intervention they need to be successful (McMillan, 2017). By identifying areas of strengths and weaknesses, diagnostic assessments can help educators and healthcare professionals develop effective intervention plans that are tailored to the individual's specific needs (Miciak *et al.*, 2020). To ensure that diagnostic assessment meets its goal, diagnostic tests must go through the crucible of appropriate test construction skills. Diagnostic assessment is also used at the workplace to reveal and intervene on work related challenges, and for employees' selection.

Employment assessment refers to the process of identifying and choosing the most qualified candidates for a job in an organisation. Effective employment selection is critical to the success of organisation as it ensures that the organisation hires individuals who possess the requisite skills, knowledge, and abilities necessary to perform the job effectively. There are several steps involved in the employment selection process, including job analysis, sourcing candidates, screening résumés, conducting interviews (which are all functions a

good test), and making job offers. Each step of the process is important and requires careful assessment to ensure that the organization hires the right candidate for the job (Scherbaum & Goldstein, 2018). Such careful assessment necessitates well-constructed tests.

One of the key challenges in employment selection is ensuring that the process is fair and unbiased. This is particularly important in the light of the growing awareness of diversity, equity, and inclusion issues in the workplace. To address these concerns, many organizations adopt new approaches to employment selection, such as blind hiring, where candidate information such as name, gender, and ethnicity are removed from résumés to reduce the risk of bias. Another important consideration in employment selection is the use of technology. With the growing availability of applicant tracking systems and other software tools, organisations are able to streamline the recruitment process and more effectively manage candidates' data. However, it is important to ensure that technology is used in a responsible and ethical manner, and that it does not exacerbate existing biases or discrimination (Deryugina *et al.*, 2021). This is why appropriate tests are needed for employment selection and other purposes. By adopting appropriate test construction practices, organizations can ensure that they apply tests that enable them hire the most qualified candidates for the job while promoting diversity, equity, and inclusion in the workplace (Briscoe & Claus, 2018).

Test Construction

According to Furr (2011), test construction refers to the process of developing and creating tests or assessments that accurately measure particular constructs or traits. It involves developing test items that are reliable, valid, and fair, and that effectively measure the knowledge, skills, abilities, or other characteristics of interest. Test construction is essential in many fields, including education, psychology, and employment, where assessments are used to make important decisions about individuals and/or groups (Osterlind, 2017). The process of test construction typically involves several key steps. These include defining the construct or trait to be measured, developing a test plan,

creating and selecting test items, administering the test, and analysing the results (Cohen & Swerdlik, 2018). Each of these steps is critical in ensuring that the test is effective and produces accurate results. The test construction process starts with the definition of constructs or traits to be measured. This involves identifying the specific knowledge, skills, or abilities that the test is intended to assess. Next, a test plan is developed that outlines the format and structure of the test. This might include decisions about the number and type of test items, the time limit for the test, and the scoring system to be used. Once the test plan is in place, test items can be developed and selected. Test items can take many forms of multiple-choice questions, essay questions, and performance tasks (Geisinger, 2013). It is important that test items are valid and reliable, meaning that they accurately measure the intended construct and produce consistent results over time (Rodriguez, 2019).

Test administration which involves giving the test to testees (test takers) to complete and ensuring that test conditions are consistent and fair for all participants. This might include providing clear instructions, ensuring that all necessary materials are available, and minimizing distractions. Then the test is scored and finally, the results of the test are analysed to determine whether they accurately measure the intended construct and whether the test is reliable and valid. Test results might be used to make decisions about an individuals' performance, such as in educational or employment assessment and appraisal, or evaluation of the effectiveness of a programme or intervention (Schmeiser & Welch, 2019). For the decision made from test results to stand the test of time, the test must go through effective test construction.

Effective test construction skills are essential for accurately measuring students' learning outcomes and providing meaningful feedback to students. To ensure that tests are well-constructed, reliable, and valid measures of students' learning outcomes, many teachers use a systematic approach to test development, such as the test blueprint. Test blueprint is a valuable tool for guiding the development of assessments by specifying the intended learning objectives, cognitive levels, and content

coverage of the assessment. By using a test blueprint, teachers can ensure that their assessments are closely aligned with learning objectives, provide appropriate test item selection, and provide a balanced coverage of content, which result in more accurate measurements of students' learning outcomes and more useful feedback to students (Salkind, 2012). Akinsola *et al.*, (2017) agreed that test construction skills refer to the ability to develop and create high-quality assessments tools that accurately measure students' learning outcomes. These skills encompass a range of tasks, such as identifying learning objectives, selecting appropriate test items, designing test blueprints, and developing scoring rubrics. Effective test construction skills require knowledge of test theory, statistical analysis, and the principles of measurement and assessment. Ajayi (2019) revealed that to develop test construction skills, individuals can engage in training programmes or coursework, such as classes in educational measurement and evaluation, psychometrics, or item response theory. Additionally, they can learn from established best practices and research on effective test construction.

Aim of Test Construction Skills

The aim of test construction skills is to enable individuals develop and design effective assessments that accurately measure the knowledge, skills, and abilities of test-takers. Test construction skills encompass a wide range of competencies, including the ability to:

1. Develop test blueprints that accurately reflect the content and skills being measured.
2. Write clear and concise test items that are free from bias and measure what they intend to measure.
3. Use appropriate item formats (e.g., multiple choice, essay, performance) based on the learning objectives and cognitive levels being assessed.
4. Determine appropriate item difficulty and discrimination levels to ensure that the test is both challenging and reliable.
5. Use statistical analysis to evaluate the psychometric properties of the test and

make appropriate revisions (Bakare, 2018).

In addition to these basic competencies, test construction skills require a deep understanding of the principles of measurement theory and an awareness of the ethical considerations involved in testing. Test developers must also be able to effectively communicate test results to various stakeholders, including educators, policymakers, and the general public (Cavanagh & Chenoweth, 2021).

Components of Test Construction Skills

Test construction skills are essential for educators, researchers, and psychologists involved in developing and evaluating tests. The components of test construction skills according to Downing and Haladyna (2019), include test design, item writing, item analysis, and test validation and reliability. Test design involves the development of test blueprints, which specify the content and format of the test. A well-designed test blueprint ensures that the test measures the intended construct and is appropriate for the target population. The blueprint outlines the number and types of items that should be included in the test, the cognitive complexity of the items, the format of the items, and the scoring criteria. Item writing commences after an appropriate test design. Item writing is the process of developing test items that measure the construct of interest. A good test item should be valid, reliable, and relevant to the construct being measured. It must adequately cover the prescribed content of programme of instruction. It is at this point that the test blueprint becomes handy. Item writers should also consider the clarity of the items, relevance of the test to the target population, and the format of the test items to be used. Some common item formats include multiple-choice, true/false, matching, arrangement, and essay. Item analysis follows after the items have been written.

Bello (2017) affirmed that item analysis is the process of evaluating the performance of individual test items. This involves examining the item difficulty and discrimination, which measure the item's ability to differentiate between high and low performers. Item analysis helps to identify items that may be too easy, too difficult, or

irrelevant to the construct being measured. Item difficulty refers to the level of challenge presented by a test item or question. It is commonly measured using various statistical techniques, such as item response theory (IRT) models, which estimate the probability of a correct response for each item based on the examinee's ability level. The difficulty of an item can have important implications for test development and interpretation of results. For instance, if a test contains mostly difficult items, it may be too challenging for the intended population and may not yield useful information. In my experience, I've found that conducting item analysis not only helps in evaluating the effectiveness of individual test items but also in refining assessments to better align with learning objectives. By examining item difficulty and discrimination, I've been able to pinpoint areas where questions may need adjustment to better assess student understanding. This process isn't just about numbers; it's about ensuring that assessments accurately measure what they intend to, ultimately leading to more meaningful insights into student performance and learning outcomes (Baruwa, 2020). On the other hand, if a test contains mostly easy items, it may not be sensitive enough to differentiate between individuals with different levels of ability (Eze, 2018).

Discrimination refers to the degree to which a test item can distinguish between individuals with high and low levels of the construct being measured. In other words, it measures how well an item can discriminate between those who perform well on the test as a whole and those who do not. Item discrimination is commonly measured using various statistical such as point-biserial correlation, which examines the correlation between an item score and the total test score. The item discrimination index is an important property of a test item, as it indicates how well the item is performing in terms of its ability to discriminate between individuals with high and low levels of the construct. Item is effective at differentiating between high- and low-performing individuals, whereas a low discrimination index indicates that the item is not effective at doing so (Smith, 2023). Gbadegesin (2019) agreed that test validation involves the process of establishing the validity and reliability of the test. Validity refers to the degree to which a test measures what it is intended to measure.

Reliability refers to the consistency of the test results over time and across different administrations. Test validation also involves examining the test's sensitivity and specificity, which measure the test's ability to accurately identify individuals with the construct being measured. Validation is the process of assessing the accuracy and reliability of something, such as a theory, model, or measurement. When using new references to validate something, it means comparing it to previously established or accepted references to determine its validity. For example, in the field of scientific research, a new hypothesis or theory may be validated by comparing its predictions to the results of previous studies; similarly, in software development, a new version of a programme may be validated by testing it against a set of known inputs and expected outputs (Yang & Chen, 2023). Reliability refers to the consistency, stability, and reproducibility of a measure or instrument over time and across different contexts. A reliable measure or instrument should produce consistent results when used repeatedly, and should not be influenced by external factors such as changes in the environment or the presence of different raters (Nguyen & Nguyen, 2023). Reliability is an important aspect of research design and is often assessed through various statistical measures such as Cronbach's alpha, test-retest reliability, inter-rater reliability, and split-half reliability (Kim & Park, 2024).

Test Blueprint

A test blueprint is a plan that outlines the content and structure of a test. It includes information about the number of questions, the type of questions, the weight of each section, and the content area that will be covered. The purpose of a test blueprint is to ensure that a test is fair, reliable, and accurately measures what it is intended to measure (Ford & Kozlowski, 2021). A test blueprint, also known as a table of specification, is a planning document that outlines the structure, content, and weighting of test items to assess students' learning outcomes; it serves as a framework for developing and evaluating tests, and helps ensure that test items are aligned with instructional objectives and standards; and provides teachers with a clear roadmap for constructing assessments that are

valid, reliable, and comprehensive (Gonzalez, 2022). The use of test blueprints has been shown to enhance the quality and validity of assessments in educational settings as Lin and Linn (2022) found that test blueprints were useful for enhancing the validity and reliability of high-stakes tests in healthcare education. In addition, test blueprints are used to improve the quality of standardized tests, as well as to ensure that tests are fair and unbiased for all students. In summary, test blueprints play a critical role in test construction for teachers: By providing a framework for planning and designing assessments, teachers can ensure that their tests accurately measure students' learning outcomes and are aligned with instructional objectives.

Well-constructed tests are critical for ensuring that tests provide valid, reliable, accurate and informative feedback on students' learning outcomes, while poorly constructed tests can lead to invalid and unreliable results (Luecht & Nungester, 2018). Effective test construction skills are important for creating assessments that are fair and equitable for all students; test items should be culturally and linguistically sensitive, free from gender or racial bias, and appropriate for students with disabilities (Hambleton & Zenisky, 2003). A test blueprint is key for teachers to develop tests that ensure that assessments are aligned with the intended learning objectives, contain appropriate test items, and provide a fair and reliable measure of students' learning outcomes (Linn *et al.*, 2017). Educational Testing Service (2018) noted that some ways in which teachers can use test blueprints to improve their test construction:

1. Identify learning objectives: Before constructing a test, teachers should first identify the specific learning objectives that they want to assess. These objectives can be organized into categories and subcategories that will form the basis of the test blueprint.
2. Determine test item types: Once the learning objectives have been identified, teachers can select appropriate test item types (e.g. multiple-choice, short-answer, essay) that align with the intended level of student learning.
3. Specify cognitive levels and weighting: The test blueprint can then be used to

specify the cognitive levels and weighting of each test item type. This can help ensure that the test is balanced and that each learning objective is assessed appropriately.

4. Develop scoring rubrics: Finally, teachers can use the test blueprint to develop scoring rubrics for each test item type. This can help ensure that the test is scored consistently and that students receive appropriate feedback on their performance (National Council on Measurement in Education, 2017).

Using a test blueprint can help teachers create assessments that are aligned with the intended learning objectives and provide a fair and reliable measure of students' learning outcomes. By following a systematic approach to test construction, teachers can improve the quality of their assessments and provide more accurate feedback on students' performance (Amrein-Beardsley & Collins, 2021).

Shepard (2000) revealed that the objectives of a test blueprint may vary depending on the type of assessment, but generally, the primary aim of a test blueprint is to provide a comprehensive and organized plan for developing a valid and reliable test. The objectives of a test blueprint include:

1. Content coverage: Test blueprint should outline the content areas and skills that the assessment will measure. This ensures that the test is comprehensive and covers all relevant areas of the subject matter.
2. Alignment with learning objectives: Test blueprint should align with the learning objectives or outcomes for the course or programme. This helps ensure that assessment is measuring what students are expected to learn.
3. Balanced representation of content: Test blueprint ensures that each content area is adequately represented in the assessment. This helps ensure that no single content area dominates the test and that all content areas are given appropriate weight.
4. Clarity and specificity: Test with specify or content areas and skills to be assessed are enhanced by the test blueprint. This

helps ensure that test developers and teachers have shared understanding of what will be tested.

5. **Validity and reliability:** Validity and reliability of a test are enhanced by the use of the test blueprint. This means that the assessment measures what it intends to measure and produces consistent results over time.
6. **Accessibility and fairness:** Test blueprint should ensure that the test is accessible and fair to all students regardless of background or circumstances. It ensures that the assessment is not at a disadvantage to any particular student or group of students (Guskey & Bailey, 2010).

Components of Test Blueprint

Nitko and Brookhart (2011) stated that a test blueprint is made up of four main components, which are content, cognitive level, number of items and item percentage or weight.

1. **Content Area:** In test construction, content area refers to the specific topics or subject matter that a test is designed to measure. Developing a clear understanding of the content area is crucial to creating a valid and reliable test. The process of identifying the content area for a test usually begins with a comprehensive analysis of the subject matter or topic to be tested. This analysis helps to identify the key concepts, skills, and knowledge that the test should elicit. Once the content area has been identified, test developers can use various strategies to create test items that effectively assess content area (Popham, 2018).
2. **Cognitive Levels:** Cognitive levels refer to the different levels of thinking that individuals use to process information and solve problems. In test construction, cognitive levels are used to describe the depth and complexity of the mental processes required to answer test items. There are several cognitive levels that can be used in test construction, including knowledge, comprehension, application, analysis, synthesis, and evaluation. These

levels are hierarchical, with each level building upon the previous level and requiring increasingly complex mental processes. Developing test items at different cognitive levels can help to ensure that a test measures a wide range of knowledge and skills. Test developers must ensure that the cognitive levels of test items align with intended learning outcomes and accurately assess the knowledge and skills being measured (Haladyna *et al.*, 2002).

3. **Number of Questions:** The number of questions in test construction is an important consideration that can affect the validity and reliability of the test results. Generally, the number of questions should be sufficient to provide a representative sample of the knowledge, skills, or abilities being assessed. For shorter tests, such as quizzes or knowledge checks, a smaller number of questions may be appropriate, such as 5-10 questions. However, for longer tests, such as exams or standardized tests, a larger number of questions is necessary to adequately assess a wide range of knowledge or skills. The number of questions should also take into account the time constraints of the test taker. If the test is too long, test takers may experience fatigue or lose focus, potentially affecting the validity of the results. In addition to the number of questions, test constructors should also consider the type of questions used, such as multiple-choice, short answer, or essay questions, and how these questions align with the learning objectives or outcomes being assessed. Ultimately, the goal should be to construct a test that is fair, valid, reliable, and aligned with the learning objectives (Sireci & Zenisky, 2006).
4. **Percentage of Items:** In test construction, the percentage of items refers to the weight or value assigned to each question or item in the test. The percentage of items is usually based on the difficulty level, the importance of the content, or the learning outcomes the test is designed to assess. The percentage of the test can be

expressed in different ways, such as the number of points or marks assigned to each item or the percentage of the total test score that each item is worth. For example, a multiple-choice test with 50 questions, each worth two points, would have a total score of 100 points. The percentage of the test can also be used to ensure that the test covers all the content

areas or topics that it is intended to assess. In this case, the percentage of the test is based on the proportion of questions or items that address each content area or topic. For example, if a test is designed to assess both knowledge and skills, the percentage of the test may be divided between knowledge-based questions and skill-based questions (Downing, 2006).

Tabular Representation of Test Blueprint of Educational Measurement and Evaluation Objective Test

Table 1: Hypothetical Test Blueprint of Test Construction in Measurement and Evaluation

Content	Behaviour/Objectives						Row Total
	Knowledge (10%)	Comprehension (10%)	Application (30%)	Analysis (20%)	Synthesis (15%)	Evaluation (15%)	
Basic Concepts Test Construction (10%)	1		2	1	1	1	6
Types of Test (10%)		1	2	1	1	1	6
Qualities of a good Test (10%)	1		2	1	1	1	6
Test Construction (20%)	1	1	4	3	2	1	12
Item Analysis and Selection (20%)	2	2	2	2	2	2	12
Test Administration scoring (20%)	1	1	4	3	1	2	12
Test Scoring (10%)		1	2	1	1	1	6
Column Total	6	6	18	12	9	9	60

Challenges of Test Blueprint

The challenges that can affect blueprinting according to Lane *et al.*, (2016) include:

1. Defining the construct being measured: It can be difficult to define the construct being measured by a test, particularly when it is a complex or abstract concept. This can lead to ambiguity in the test blueprint and make it difficult to ensure that the test is measuring the intended construct.
2. Balancing content areas: Creating a balanced test blueprint that accurately reflects the content areas and cognitive levels being assessed can be a challenge. It can be difficult to determine how much weight to give to each content area and

how to allocate the available testing time to each area.

3. Aligning with standards: Ensuring that the test blueprint is aligned with relevant educational standards and learning objectives can be a challenge. This requires a thorough understanding of the relevant standards and how they relate to the content being assessed.
4. Ensuring fairness and equity: Test blueprinting can present challenges in ensuring that the test is fair and equitable for all test-takers, particularly when it comes to accommodating diverse cultural and linguistic backgrounds.
5. Keeping up with changing needs: Test blueprints need to be regularly reviewed

and updated to reflect changes in curriculum, educational standards, and the needs of the students being assessed.

Relevance of Test Blueprints for Effective Teacher Test Construction Skills

The use of test blueprint is highly relevant for effective teachers' test construction skills and quality assessment for several reasons which include the following:

1. **Alignment with learning objectives:** One of the primary benefits of using test blueprints is that they facilitate alignment with learning objectives. Test blueprints provide a detailed overview of the content that students are expected to master, and this information can be used to construct test items that align with specific learning objectives. For example, a test blueprint may indicate that students are expected to understand the principles of algebra, including linear equations, quadratic equations, and graphing. A teacher can use this information to construct test items that assess students' knowledge of these specific algebraic concepts (Nitko, 2016).
2. **Appropriate test item selection:** Another important aspect of test construction is the selection of appropriate test items. Test blueprint helps teachers choose appropriate test items by providing information on the specific types of questions that are most effective in assessing different learning objectives. For example, a test blueprint may indicate that multiple-choice questions are most effective in assessing students' knowledge of algebraic equations, while open-ended questions are more effective in assessing students' ability to apply algebraic principles to real-world situations (Downing, 2006).
3. **Balanced coverage of content:** Test blueprint ensures that tests have a balanced coverage of content. This is important because it ensures that all topics and concepts are assessed, and that no topics are overrepresented or underrepresented. Test blueprint helps teachers achieve this balance by providing a clear overview of the content that should be covered on the test, and the percentage of questions that should be devoted to each topic. For example, a test blueprint may indicate that 20% of the test should assess students' knowledge of linear equations, 30% should assess quadratic equations, and 50% should assess graphing (Haladyna & Downing, 2004).
4. **Improved reliability and validity:** Reliability and validity are two important aspects of test quality. Reliability refers to the consistency and stability of test scores over time, while validity refers to the extent to which a test measures what it is intended to measure. Test blueprints improve both the reliability and validity of tests by ensuring that all test items are directly aligned with specific learning objectives, and that all topics and concepts are assessed based on assigned weights. This ensures that all students are evaluated using the same standards, and that test scores accurately reflect students' knowledge and skills (Popham, 2006).
5. **Improved Test Quality:** Test blueprints improve the overall quality of tests. By providing a clear overview of the content covered on the test, test blueprints ensure that all relevant topics and concepts are included, and that no important areas are overlooked. This helps ensure that tests accurately assess students' knowledge and skills, and that test scores provide meaningful information that can be used to guide instruction and improve students' learning outcomes (Kane, 2006).
6. **Increased Efficiency:** Test blueprints increase the efficiency of test construction. By providing a clear framework for test construction, test blueprints save teachers time and effort by streamlining the test development process. Instead of starting from scratch each time a test is needed, teachers use test blueprints as a guide, allowing them to quickly and easily develop tests that are aligned with specific learning

objectives and provide balanced coverage of content (Hambleton & Zenisky, 2011).

7. **Increased Consistency:** Another benefit of using test blueprints is that they promote consistency across multiple tests. By using the same blueprint for multiple tests, teachers can ensure that all tests are constructed using the same standards and guidelines, ensuring consistency in the evaluation of students' learning. This consistency is important because it ensures that students are evaluated using the same criteria, regardless of the teacher or course (Haladyna & Downing, 2004).
8. **Improved Alignment with Curriculum:** Test blueprints improve the alignment of tests with the curriculum. By providing a clear overview of the content covered on the test, test blueprints ensure that all tests are directly aligned with the curriculum and the specific learning objectives that have been established. This alignment is important because it ensures that tests accurately reflect the content and skills that students are expected to master, and that test scores provide meaningful information that can be used to guide instruction and improve students' learning outcomes (Wiggins, 1998).

Conclusion

The use of test blueprints is important and effective for improvement of teachers' test construction skills. Test blueprints provide systematic approaches to test development, ensuring that tests are well-constructed, reliable, and valid measures of students' learning outcomes. By using a test blueprint, teachers can ensure that their assessments are closely aligned with learning objectives, provide appropriate test item selection, and provide a balanced coverage of content. The use of test blueprint enables teachers' create assessments that provide accurate measures of students' learning outcomes and useful feedback to students. The use of test blueprints is a relevant and valuable approach to effective teacher test construction. By providing a systematic approach to test development, teachers

can improve the reliability, validity, and fairness of their assessments, which ultimately results in more accurate measurements of students' learning outcomes.

Suggestions

Here are some suggestions on the use of test blueprint in improving teachers' test construction skills for quality assessment:

1. Teachers should have a clear understanding of test blueprint which outlines the content, objectives, and skills that a test should assess.
2. Teachers should use test blueprint as a guide for developing test items that will accurately assess the intended knowledge and skills.
3. Teachers should improve their test construction skills by developing high-quality test items using test blueprint.
4. Teachers should improve on feedback to students on their performance using test blueprint.
5. Teachers should continuously evaluate the test blueprint to ensure that it is still relevant and valid and make updates when necessary.
6. Teachers should develop test blueprint to ensure accurate reflection of the content and skills being measured.
7. Teachers should write clear and concise test items that are free from bias and measure what they intend to measure.
8. Teachers should use appropriate item formats (e.g., multiple choice, essay, performance) based on the learning objectives and cognitive levels being assessed.
9. Teachers should determine appropriate item difficulty and discrimination levels of tests to ensure that the tests are both challenging and reliable.
10. Teachers should use statistical analysis to evaluate the psychometric properties of tests and make appropriate revisions when necessary.

References

- Ajayi, I. A. (2019). Evaluation of test construction skills among Nigerian teachers: A comparative Study. *African Journal of Educational Research*, 11(2), 45-56.
- Akinsola, M.K., Tella, A. & Tella, A. (2017). Assessment of test construction skills among teachers in Nigeria: Implications for quality education. *Journal of Education and Practice*, 8(4), 11-18.
- Allen, J. J., & Yen, W. M. (2021). Validity and reliability of a new computer-adaptive test for assessing English language proficiency. *Language Testing*, 38(1), 48-69.
- American Psychological Association. (2020). *Publication manual of the American psychological association* (7th ed.). American Psychological Association.
- Amin, T., Shabbir, M., & Amin, N. (2021). Writing better multiple choice questions: A guide for novice test writers. *Medical Education Online*, 26(1), 50-58.
- Amrein-Beardsley, A., & Collins, C. (2021). Beyond test blueprints: Using frameworks to enhance teacher-created assessments. *Educational Assessment, Evaluation and Accountability*, 33(1), 1-22.
- Bakare, J. A. (2018). Enhancing test construction skills: A case study of Nigerian teachers. *International Journal of Educational Psychology and Counseling*, 2(1), 22-33.
- Baruwa, A. O. (2020). Investigating the test construction skills of Nigerian teachers: A mixed-methods approach. *Journal of Educational Measurement*, 15(3), 78-89.
- Baumeister, R. F., & Vohs, K. D. (2022). Test blueprints and the measurement of self-control. *Journal of Personality Assessment*, 104(1), 25-36.
- Bello, S. O. (2017). Assessing test construction skills among Nigerian educators: A comparative analysis. *Nigerian Journal of Educational Assessment*, 5(2), 33-44.
- Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice*, 25(6), 551-575.
- Briscoe, J. P., & Claus, L. (2018). *Handbook of employee selection*. London: Penguin Random House.
- Cavanagh, M., & Chenoweth, L. (2021). Constructing effective multiple-choice questions to measure learning outcomes in online environments. *Journal of Applied Research in Higher Education*, 13(1), 94-107.
- Chen, C. H., & Chen, H. L. (2022). The effects of online quizzes on learning: A meta-analysis. *Computers & Education*, 174, 104-129.
- Cohen, R. J., & Swerdlik, M. E. (2018). *Psychological testing and assessment: An introduction to tests and measurement* (9th ed.). McGraw-Hill Education.
- Deryugina, T., Shurchkov, O., & Babajanian, B. (2021). Algorithmic bias in hiring: An audit study. *Journal of Political Economy*, 129(4), 1241-1284.
- Downing, S. M. (2006). Twelve steps for effective test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 3-25). Lawrence Erlbaum Associates.
- Downing, S. M. (2019). Validity and reliability of assessment in medical education. *AMEE Guide No. 37. Medical Teacher*, 41(3), 271-279.
- Educational Testing Service. (2018). Test blueprinting: An overview. Retrieved from <https://www.ets.org/>
- Eze, C. E. (2018). Enhancing test construction skills among Nigerian teachers: A professional development approach. *Journal of African Educational Research*, 7(1), 56-68.
- Fives, H., & DiDonato-Barnes, N. (2013). Classroom test construction: The power of a table of specifications. *Practical Assessment, Research, and Evaluation*, 18(3), 12-17.
- Ford, J. K., & Kozlowski, S. W. (2021). Developing and using test blueprints: A primer for practitioners. *Organizational Research Methods*, 24(1), 1-26.
- Furr, R. M. (2011). *Scale construction and psychometrics for social and personality psychology*. SAGE Publications.

- Gbadegesin, S. (2019). Test construction skills among Nigerian teachers: Challenges and solutions. *African Educational Research Journal*, 6(3), 112-125.
- Geisinger, K. F. (2013). *APA handbook of testing and assessment in psychology: Test theory and testing and assessment in industrial and organizational psychology*. American Psychological Association.
- Gonzalez, M. (2022). Test blueprint development for a computer-based test of English for academic purposes. *Language Assessment Quarterly*, 19(1), 49-71.
- Guskey, T. R., & Bailey, J. M. (2010). *Developing grading and reporting systems for student learning*. Corwin Press.
- Gutiérrez, D., García, J. I., & Llor, M. (2020). Virtual reality psychological assessment: Advantages, limitations, and future challenges. *Frontiers in Psychology*, 11, 20-26.
- Haladyna, T. M. (2019). *Developing and validating multiple-choice test items* (4th ed.). United Kingdom: British Publishing Company.
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 23(1), 17-27.
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 23(1), 17-27.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309-334.
- Hambleton, R. K., & Zenisky, A. L. (2011). Criterion-referenced testing. In M. L. Kamil, P. D. Pearson, E. B. Moje, & P. P. Afflerbach (Eds.), *Handbook of reading research*, (pp. 509-539). Routledge.
- Herman, J. L., & Golan, S. (1991). Effects of standardized testing on teachers and learning another look (CSE Technical Report 334). National Center for Research on Evaluation, Standards and Student Testing (CRESST) UCLA Graduate School of Education. <https://files.eric.ed.gov/fulltext/ED341738.pdf>
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *educational measurement* (4th ed., pp. 17-64). American Council on Education/Praeger.
- Kim, Y. J., & Park, H. S. (2024). Reliability estimation of rubric-based performance assessments: A comparative study of classical and modern test theory approaches. *Educational Assessment*, 29(1), 67-84.
- Krouska, A., Troussas, C., & Virvou, M. (2018). Computerized adaptive assessment using accumulative learning activities based on revised bloom's taxonomy. In Joint Conference on *knowledge-based software engineering* (pp. 252-258). Springer Charm
- Lane, S., Raymond, M. R., & Haladyna, T. M. (2016). *Handbook of test development*. United Kingdom: Hodder & Stoughton.
- Lee, Y. H., & Huang, W. D. (2018). Developing a test blueprint for a computerized adaptive testing system for assessing English proficiency. *Language Testing*, 35(3), 329-348.
- Lin, T. J., & Linn, M. C. (2022). Test blueprint design for innovative science assessments. *Journal of Research in Science Teaching*, 59(2), 218-238.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (Eds.). (2017). *Complex, interdisciplinary assessments: Theory and practice*. New York: Penguin Random House.
- Luecht, R. M., & Nungester, R. J. (2018). *Developing and validating test items*. In C. A. Clauser, E. F. Downing, & T. M. Haladyna (Eds.), *Handbook of test development* (2nd ed.). United Kingdom: Pluto Press
- McMillan, J. H. (2017). *Classroom assessment: Principles and practice for effective standards-based instruction*. Pearson
- Miciak, J., Horn, I. S., & Schweinle, W. (2020). The role of diagnostic assessment in evaluating teaching quality. *Review of Research in Education*, 44(1), 240-265.
- National Council on Measurement in Education. (2017). Guidelines for developing assessment blueprints. Retrieved from <https://www.ncme.org/>

- Nguyen, T. H., & Nguyen, H. T. (2023). Investigating the reliability and validity of a mathematics achievement test using Rasch analysis. *Journal of Educational Measurement, 39*(4), 412-430.
- Nitko, A. J. (2016). *Educational assessment of students* (7th ed.). Boston, MA: Pearson.
- Nitko, A. J., & Brookhart, S. M. (2011). *Educational assessment of students* (6th ed.). Boston: Pearson Education.
- Osterlind, S. J. (2017). *Constructing test items: Multiple-choice, constructed-response, performance and other formats* (4th ed.). United Kingdom: Verso Books.
- Pellegrino, J.W. (2022). Using test blueprints to improve the quality of assessments in educational settings. *Educational Assessment, Evaluation and Accountability, 34*(1), 67-83.
- Popham, W. J. (2006). *Assessment for educational leaders*. Pearson.
- Popham, W. J. (2018). *Classroom assessment: What teachers need to know* (8th ed.). Pearson Publishers Ltd.
- 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.
- Rodriguez, M. C. (2019). *Test development and validation: A primer for teachers and test developers*. United Kingdom: Atlantic Books.
- Salkind, N. J. (2012). *Tests and measurement for people who (think they) hate tests & measurement* (2nd ed.). Sage Publications.
- Scherbaum, C. A., & Goldstein, H. W. (2018). *An introduction to competency-based selection*. New York: Zed Books.
- Schmeiser, C. B., & Welch, C. J. (2019). *Principles of test development*. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 29-42). London: Quercus.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.
- Sireci, S. G., & Zenisky, A. L. (2006). Factors affecting the validity of cross-lingual assessments. *Educational Measurement: Issues and Practice, 25*(4), 14-21.
- Smith, J. K. (2022). The impact of test blueprint use on student performance and learning measurement. *Journal of Educational Assessment, 39*(1), 43-56.
- Stiggins, R. J. (2020). Assessment for learning revisited: An historical and contemporary analysis. *Assessment in Education: Principles, Policy & Practice, 27*(3), 247-266.
- Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. Jossey-Bass.
- William, D. (2018). Formative assessment: Ten years on. *British Educational Research Journal, 44*(3), 375-378.
- Yang, H., & Chen, W. (2023). Investigating the reliability and validity of a writing assessment rubric for English language learners: A mixed-methods study. *Language Assessment Quarterly, 30*(3), 291-310.