



BUILDING QUALITY ITEM BANK FOR ASSESSMENT OF STUDENTS' LEARNING OUTCOMES

OKECHUKWU, John Ndubueze; PhD¹

Department of Educational Management and Policy, Nnamdi Azikiwe University,
P.M.B 5025, Awka, Nigeria; E-mail: oj.ndubueze@unizik.edu.ng

JOSEPH, Ayodele Abiodun; PhD²

Department of Educational Management and Policy, Nnamdi Azikiwe University
P.M.B 5025, Awka, Nigeria; Email: aj.ayodele@unizik.edu.ng

IRONKWE, Athanasius Chinweze; PhD³

Holy Family Fathers and Brothers of the Youth,
P.M.B 86, Okija, Nigeria; Email: athanchis@yahoo.com

MANAFA, Ngozi Florence; PhD⁴

Department of Educational Foundations, Chukwuemeka Odumegwu Ojukwu University P.M.B
6059, Igbariam Campus, Nigeria; Email: ngomanafa@gmail.com

Abstract

This opinion paper critically examines the process of building a quality item bank for the assessment of students' learning outcomes. Grounded in the Classical Test Theory (CTT) as propounded by Charles Spearman (1904) and expanded by DeMars (2018). The paper explores the concept and core functions of an item bank, emphasizing its role in promoting standardized, valid, and reliable assessments. It outlines the systematic phases of item bank development, beginning with effective planning and progressing through the construction of quality test items, development of test blueprints or tables of specification, and application of rigorous technical criteria. The paper highlights the significance of item writing guided by Bloom's taxonomy, as well as the critical roles of expert review, panel moderation, editing, and trial testing. Furthermore, it emphasizes item analysis as essential for evaluating item functionality and refining test quality before final item banking. The paper advocates for continuous quality assurance mechanisms to enhance test fairness and measurement precision. In conclusion, it suggested strategic recommendations for integrating digital platforms, expert collaboration, and psychometric standards in item bank development. This structured approach ensures that assessments not only measure learning outcomes accurately but also support instructional improvement and policy decisions.

Keywords: Item Bank; Building Quality Item Bank; Assessment; Students' Learning Outcomes

Introduction

Assessment is a central element in the overall quality of teaching and learning in higher education. Therefore, teaching and learning process is incomplete without effective and



meaningful assessment of students learning outcomes. A well-designed assessment sets clear expectations, establishes a reasonable workload for both staff and students and provides opportunities for students to self-monitor, rehearse, practice and receive feedback (Brown & Harris, 2021). Learners' assessments are used to identify progress that indicates what has been achieved and what needs to be done by providing objective measurement to serve as reference point for future work. Assessments provide opportunities to ask tutors questions, raise content issues that illuminate expected standards for assignments, and provide information or additional resources to address weak areas or to pursue topics in greater depth (AERA et al., 2018). For these attributes of assessment to be realized in higher education especially in the open and distance learning, there must be a well designed and developed item bank to enable the conduct of meaningful assessment. This is very relevant in open distance learning environment especially now that open and distance learning all over the world is technologically driven and the recent trend is towards its use in assessment of learning outcomes (Chikwere & Opara, 2023).

Besides, the peculiar nature of Open and Distance Learning (ODL) institutions, administrative costs and security, and automation of the testing process has become an issue of great importance in such systems. Automated item banking and quality test development will enhance assessment of learning outcomes in the ODL institutions and in other institutions desiring to automate their assessment process or are battling to contend with the large number of examinees. For instance, the use of technology in assessment as is currently practice by the premier ODL institution in Nigeria underscores the urgent need for handy item bank to strengthen the existing practice (Adebayo & Adetunji, 2020). Thus, a robust repository of test items and the components that make up those items, are critical to keeping examination content secure and fresh (Ajayi & Lawal, 2019). Test items constructed from such item bank will be very useful once the item bank has a large collection of test items in each cell of the assessment blueprint stored in computer storage together with their parameters. Item bank is very relevant to ODL scenario because all the tutors are not always involved in assignment design. Assessment design is often undertaken by course coordinators and sometimes by course writers based on pre-determined criteria for the assessment.

Assessment practice includes assessment design and assessment strategies. Morgan and O'Reilly (as cited in Bello & Yusuf, 2019) identified the following as elements of good ODL assignment design:

- Clear rationale and consistent pedagogical approach;
- Explicit values, aims, standards and criteria;
- Authentic and holistic tasks;
- Facilitative degree of structure (the right balance between direction and flexibility);
- Sufficient and timely formative assessment; and
- Awareness of the learning context and perceptions.

McMillan (2018) observed that the features of assignment design and assessment strategies that support learning and achieve the goals of assessment should be fair, clear, central, appropriately challenging, interesting, responsive, useful, effective, flexible and supportive of further learning. It is therefore helpful to develop and use effective assessment items. This can only be achieved by utilizing well-planned assessments for a course whether or not course tutors are involved in the planning of assessment. Hence, the need for a well-developed item bank to promote the principles of setting efficient assessments practice with the intent of achieving all the complementary attributes associated with ideal practice. Item bank fits well into the open and distance learning mode of instructional delivery and assessment practices due to the collaborative

way of doing things, with openly sharing and learning together. A rich and quality item bank will immensely help to achieve the goals of the ODL institutions assessments scheme as a first step toward competent management and quality control; since ODL institutions, notable for embracing large class sizes, are the emerging mode of transformational approach to education in Nigeria and globally.

Quality item bank has the potential to support knowledge management; communities of practice and project delivery characterizing ODL mode of instructional delivery; once the item bank is made easily accessible to all who might use the bank by establishing it in a central locale. A good item bank possessing the right quality of items is an asset to assessment of students learning outcomes and a worthwhile investment in learning in open and distance learning scenarios (Eze & Mbanusi, 2018). This is because ODL institutions are characterized by large classes with students dispersed in various locations and they are still assessed centrally for their certifications and award of degrees. To develop item bank, subject matter experts first create the blueprint for questions sets by mapping the items to the assessment's requirement. Item writers then create items as per the approved blueprint and follow the guidelines provided for instructional strategies. Items may be supported by relevant images, animations, audio or videos to make them more interactive (Tao & Chang, 2020). Hence, this paper discusses the process of development of quality item bank for effective and responsive assessment of students' learning outcomes under the following subtopics: concept and functions of item bank, planning item bank, the test blueprint; item writing, item review process, and item banking. The development of quality item bank depends on the premise that by developing a bank of well-written items, course coordinators can effectively construct and manage both formative and summative assessments. This can be done by creating assessments that accurately measure knowledge, skills and abilities necessary for competence. A standard item banking system not only provides a set of tools to facilitate the writing, review, editing and selection of test questions, but also provides the automation, standardization and scalability essential to developing and maintaining effective responsive assessment.

Theoretical Framework

This theoretical study was anchored on the Classical Test Theory (CTT). Classical Test Theory (CTT) was primarily developed by Charles Spearman in 1904 and advanced by DeMars in 2018. CTT, rooted in psychometric analysis, posits that an observed test score (X) is the sum of a true score (T) and random error (E), i.e., $X = T + E$. It assumes that each test item contributes to the total score and that the reliability of a test is enhanced by minimizing errors and ensuring internal consistency. CTT provides a foundational basis for evaluating test reliability, item difficulty, and discrimination indices, which are core principles in item bank development. The theory underscores the need for consistency in test construction by evaluating the psychometric properties of items across different student populations. Although CTT has limitations such as its dependence on the sample used it remains an accessible and widely applicable framework for educational assessments, particularly in resource-constrained educational systems like those in many developing countries.

In the context of building a quality item bank, CTT is particularly relevant as it guides the systematic development, calibration, and validation of test items to ensure they reliably assess students' learning outcomes. It provides a methodological pathway for selecting and retaining only those items that demonstrate strong psychometric properties such as high reliability coefficients and strong item-total correlations. Through the lens of CTT, educational researchers and assessment developers can construct item banks that are fair, valid, and capable of capturing the true achievement levels of diverse learners. Moreover, CTT facilitates item analysis procedures

such as computing item difficulty and discrimination indices; which help ensure the balance of easy, moderate, and difficult items in the bank. Its relevance extends to policy and practice, supporting educational accountability and instructional improvement through valid assessments.

Concept and Functions of Item Bank

An item bank in education is a curated repository of validated assessment questions designed to measure learning outcomes consistently across time, settings, or populations. It supports efficient, flexible, and psychometrically sound test development. Usman and Musa (2019) described item bank as a planned library of test items pooled through comparative efforts under protection and support of an institution for the use of evaluators, academics and students in partial fulfillment of the requirements of the teaching learning process (Camilli et al., 2017). Brown and Harris (2021) referred to item bank as a systematically organized collection of test questions (items) that are stored electronically or manually and are used for assessing students' knowledge, skills, attitudes, or performance. Each item in the bank is typically classified by subject area, learning objectives, cognitive level (e.g., Bloom's taxonomy), difficulty level, format (e.g., multiple choice, essay), and statistical properties (e.g., item difficulty, discrimination index, reliability). Basically, item bank is a repository or database of questions from which one can draw on. Test items in this paper refer to tasks and/or items that are used to construct tests and examinations. This is so defined because items are not necessarily in question format and are sometimes independent of one another. On the other hand, the range of techniques involved in preparing the items is referred to as item writing. It draws attention to the separate skills, abilities or pieces of knowledge which make up human learning. The test item writers consider what happens when these items achieve significance or importance by having been grouped or combined as building blocks with other items to form a test instrument, the larger whole (Bature & Sunday, 2020). Items are developed, assembled, collected, organized, reviewed, coded, used and refined in ways that make sense by providing value to the organization. Item bank is therefore, a term generally used for a store of items which enables effective use and timely response to intended future assessment of students learning outcomes. It stores items independently of their development for any particular assessment. Ideally, the item bank is where new items are created, existing items edited and previewed, and where most assessment processes are based (Liu & Yin, 2019).

It is important to differentiate item creation from test creation. Item bank treat items as resources which can be collected, stored and rolled out one or more times to be combined into different tests. A module can have item bank with thousands' (1000s') of items, but only a hundred (100) or so might actually be live at any one time, depending on the current assessment (or test) definitions. Alternatively, item bank might have only about ten (10) items, but reusing them might result in quizzes with perhaps fifty (50) or hundred (100) items in total. Therefore, items in item bank are thought of as a long-term resource. Item bank is thus a database of items that can be shared between and among courses, learning object repositories, programmes and courses and is pre-designed to enable items meeting specific criteria to be drawn from them to create assessments (Alade & Ojo, 2021). An assessment practice is given momentum by quality items. Good items represent an investment in learning. Item bank helps in better management of that investment. But, coming up with a good item is part art, part science. It is therefore a good practice to assemble good items in item bank where it could be shared, used and learn from.

Functions of Item Bank

The chief function of item bank is to enable users to locate particular test items housed in the bank for use in assessment of learning outcomes in any given area of the curriculum planned for in the development and during the establishment of the bank. Ideally, item bank resource is designed to be searchable databank of items from various programmes and courses offered by the institution. Item bank can support knowledge management by providing a medium for retaining and passing on knowledge such as the items that work the best from one generation to the next; communities of practice whereby item bank provides groups of people who share a common set of problems, concern or interests with a way to learn and grow as a group; and project delivery in which work teams can draw on item bank for items that help to better understand project stakeholders and their needs (Ali & Musa, 2018). These underpin the importance of call for comprehensive item bank for the various programmes and courses offered by ODL institutions bedeck by large students' population, large number of programmes and courses, resulting in the task of assessment being too cumbersome to manage (Nwankwo & Uzoechina, 2017). To enhance the practice of assessment, item bank serves three functions namely item entry and storage; item retrieval for reviewing items, formatting test forms, and editing and updating items; and maintenance of item history. These functions enhance the practice of assessment and the use of computer technology provides automated scoring and item analysis; and preparation of records and reports, including Grade book (Wilson, 2016).

A well-developed item bank has the advantage of relieving the course tutors of frequent writing of new items for every assessment. This is so because a well designed and developed item bank is much more complete in covering the domains of educational objectives once the items are constructed based on the blueprint and item development guidelines. Items can be edited and corrected periodically which makes the items in the bank to be responsive and ideal for particular assessment. The large number of items per programme per course enhances generation of equivalent forms of much better-quality items than those hurriedly written for a single assessment once the item bank is rich enough to forestall over usage of items for particular set of assessments. Because, it ideally consists of a large number of items for each course objective; which makes the generation of multiply forms subsequently possible as alternatives to earlier versions, thereby maintaining the quality of the assessment. It is also easier to put an examination together, and it enables quicker scoring and analysis. Other advantages enumerated by Kamba & Okonkwo, (2023: 120) are that:

"parallel test can be generated from the bank which would allow learners who were ill for a test or due to some other reasons were unavoidably absent to take up the test later; they are cost effective since new questions do not have to be generated at the same rate from year to year; the quality of items gradually improves with modification of the existing ones with time; and the burden of test preparation is considerably lightened when enough high quality items have been assembled in the item bank".

Planning Item Bank

The most essential step in developing item bank is planning. This involves the preparation of individuals, the identification of the rationale for starting item bank, and the identification of the aim to be accomplished with the item bank. The staff can formally be trained by using in-house personnel, bringing in a traveling workshop, or having people attend a pre-session at a research association conference to knowledgeable understand and participate in the project (Downing &

Haladyna, 2016). Also, the suggestions of Downing and Haladyna (2016) are worthwhile and are that item bank project should have:

- senior level knowledgeable personnel to answer technical questions that might arise;
- computer experts that are capable of performing relevant tasks such as modifying computer programmes, establishing a database system and are capable of running packaged programmes.

In addition, there should be:

- data processing staff who should examine literature and programmes on item banking to determine the programmes to be developed and those to modify to suit the institution's purposes and assessment needs.
- While the staff in-charge of assessment in the institution have to coordinate activities to assure that the data requirements are met.

Quality Items Construction

Students' assessment instrument such as test items serves a variety of uses regardless of the type of instrument used or how the results are to be used. All of the measurements should possess certain characteristics. The most essential of these are validity, reliability and usability (Isaacs & Gultig, 2021)). It is necessary to ensure that items to be included in the item bank possess these characteristics.

Validity: The extent to measure what it intends to measure. It has four types; the content, construct, concurrent and predictive. Test validity can be affected by: inappropriateness of the test, direction, vocabulary and construction of test, level of difficulty, constructions, length of test, arrangement of items, and patterns of answers.

Reliability: Is the consistency of scores obtained by an individual given the same test at different times. This can be estimated using test-retest method, alternate forms, split-half and kuder-Richardson Formula 21. The reliability of test may however be affected by the length of test, the difficulty of test item, the objectivity of scoring heterogeneity of the student group, and limited time.

Usability of Test Means its Practicability: This quality is determined by: ease in administration (administrability), ease in scoring (scorability), ease in interpretation and application (interpretability), economy of materials, and the proper mechanical make-up of the test. A test to be effective must be valid. A valid test is always valid, but not all reliable test is valid (Popham, 2017).

It is important to consider and to ensure that items to be pooled into the item bank possess these characteristics. Ensuring validity, reliability and usability lead to quality test construction. Generally, high quality assessments are considered those with a high level of reliability and validity. A good assessment has validity and reliability, plus the other quality attributes for a specific context and purpose. It should be noted that it is not always possible to totally package assessment that is totally valid and totally reliable in practice. But effort could be made during the process of developing items and packaging same for defined assessment to enshrine these attributes as much as possible and to ensure a balance tradeoff between reliability and validity.

Test Blueprint or Table of Specification

A test blueprint or table of specification consists of a matrix, or chart representing the number of items to be included in the test within each topic and level of objective. The blueprint identifies the objectives and skills that are to be tested and the relative weight on the test given to each. It helps to ensure that the desired coverage of topics and levels of objectives are obtained.

Once blueprint is created, items are written to match the levels of objectives within each topic. Each cell in the table corresponds to a particular task and the subject content. By specifying the number of test items for each cell, the emphasis to give each task and each content area is determined. A test, no matter how extensive contains only a sample of the many possible test items that could be included (Madaus, Russell, & Higgins, 2018)). Therefore, sample items should be as representative as possible of the various areas of focus in the blueprint for assessment. It is important to design the whole structure in detail before the items writer starts to write.

Test blueprint is economical of tutor's time. It is necessary from the point of view of sound practice, to have a thoroughly planned view of the whole exercise before one starts to assemble together the items to meet the objectives. In other words, the design of question paper which guides the blueprint considers setting percentage weight to these parameters: learning objectives; types of items; content areas in a module; and difficulty level. A blueprint designed based on the parameters serve the purpose professionally. The resulting two-dimensional tables specify the level of objectives (expected learning outcomes) in relation to the content (modules/ units) of a course. This is illustrated in Table 1. It defines how many items are to be taken from each of the contents, module/unit for each level of learning objective and for each type of item. A detailed table of specification helps to ensure representativeness of sample of item for learners' assessment. Developing and using test blueprint in item writing ensures adequate coverage of the content of the module/unit and achievement of the expected learning objectives. Once multiples of items are written for each cell of the blueprint, known item parameters enable strict control on generation of items for assessment by making it possible to search for items according to number of criteria.

Developing a Test Blueprint

As earlier stated, a test blueprint is a two-dimensional table that specifies the level of objectives in relation to the content of the course. The two dimensions (content and objectives) are put together in a table by listing the objectives in hierarchical order (say Bloom's taxonomy of cognitive educational objectives) across the top of the table (horizontally) and the content (Modules to be covered in the content) down the table (vertically) as shown in Table 1a. This provides the complete framework for the development of test items. The test blueprint is then developed by a deliberate plan designed to take care of the coverage of content and objectives in the right proportion according to the degree of relevance and emphasis (weight) attached to the various content areas (Modules) and the objectives (at the various levels) in the teaching learning process. That is, first consideration in the development of the blueprint is the weight to be assigned to higher order items and lower order items (the educational objectives hierarchical levels) across the top of the table (horizontally). That is for instance, educational objectives at the higher and at the lower cognitive levels. Same considerations are applicable to the other domains of educational objectives (psychomotor and affective domains). The above consideration is utilized in the allocation of numbers of items to be developed in each cell under content and objective dimensions.

Table 1a: A Sample of Comprehensive Assessment Blueprint on Foundations of Education (Initial Table)

Content		Objective Items							Subjective Items					
		Multiple Choice Questions							Very Short Answer Questions				Long Answer Questions	
Area	Weight (%)	Kn (10%)	Cp (15%)	Ap (15%)	An (30%)	Sv (10%)	Ev (20%)	Total (100%)	Kn (10%)	Cp (15%)	Ap (15%)	Sk (60%)	Total (100%)	D (100%)
Module 1	20													
Module 2	20													
Module 3	25													
Module 4	25													
Module 5	10													
Total	100							200					100	20

Kn = Knowledge, Cp = Comprehension, Ap = Application, An = Analysis, Sv = Synthesis, Ev = Evaluation, Sk = Skill, D = All Objectives Defined.

The hypothetical case under consideration is similar to what the National Open University of Nigeria (NOUN) should ideally be using considering the nature of her tutor marked assignments (TMAs) (Objective test items) and end of semester examinations (ESEs) (Subjective test items) or submultiples of this sample. Each of the sections (Objective Items and Subjective Items) of the comprehensive blueprint can be used in isolation depending on what the assessment practice of the institution is. In this hypothetical case, the level of difficulty for lower order items (knowledge (Kn), Comprehension (Cp) and Application (Ap)) is 40% of the total items to be developed in the objective items section while the higher order items (Analysis (An), Synthesis (Sy) and Evaluation (Ev)) is 60%. That is, 40% of the total items targets lower order while 60% targets higher order. The items are spread across the lower order and the higher order accordingly. Decisions on the number of items to be set are basis for determining items for each content area. For example, in Table 1a Module 1 is weighted 20% and 200 items are to be written in all in the Objective Items section. Therefore, total number of items for module 1 at the various levels of educational objectives is obtained thus:

Module 1, weight: 20% of 200 items = 40 items.

Likewise, numbers of items for the other Modules are computed using the same procedures. Hence for the other Modules are computed accordingly:

Module 2, weight: 20% of 200 items = 40 items;

Module 3, weight: 25% of 200 items = 50 items;

Module 4, weight: 25% of 200 items = 50 items;

Module 5, weight: 10% of 200 items = 20 items.

Total number of items in the Content dimension = (40 + 40 + 50 + 50 + 20) items = 200 items.

The computed values are then listed against each content area at the total column for the Objective Items section to correspond with its particular content in Table 1b. The same procedure is repeated for the Objective dimension. Just like in the above computations;

Knowledge (Kn), weight: 10% of 200 items = 20 items;

Comprehension (Cp), weight: 15% of 200 items = 30 items;

Application (Ap), weight:	15% of 200 items	=	30 items;
Analysis (An), weight:	30% of 200 items	=	60 items;
Synthesis (Sy), weight:	10% of 200 items	=	20 items;
Evaluation (Ev), weight:	20% of 200 items	=	40 items.

Total number of items in the Objective dimension = (20 + 30 + 30 + 60 + 20 + 40) items = 200 items.

In the above cases also the worked-out values are listed against each objective at the last horizontal row, alongside the provision for total.

Finally, the items for each cell are calculated and distributed to the various levels of objectives accordingly following the assigned weights and numbers of items required at both the Objectives and the Contents dimensions. For instance, for Module 1;

Knowledge (Kn), weight:	10% of 40 items	=	4 items;
Comprehension (Cp), weight:	15% of 40 items	=	6 items;
Application (Ap), weight:	15% of 40 items	=	6 items;
Analysis (An), weight:	30% of 40 items	=	12 items;
Synthesis (Sy), weight:	10% of 40 items	=	4 items;
Evaluation (Ev), weight:	20% of 40 items	=	8 items.

Total number of items in the Objective dimension of Module 1 = (4 + 6 + 6 + 12 + 4 + 8) items = 40 items.

The same process is repeated for Module 2 to Module 5 and the expected actual number of items are calculated and written against the appropriate cells step by step until all the cells are filled-up (Table Ib). The overall procedure is repeated for the Subjective Items Section of the test blueprint. The test blueprint now completed serves as guide for item writing. Once the test blueprint is developed, item writing would then commence. Items written based on the developed test blueprint would normally have appropriate content validity at the required level of difficulty as long as the appropriate procedures explained above are adhered to while developing the test blueprint for item writing.

Table Ib: A Sample of Comprehensive Assessment Blueprint on Foundations of Education (Final Table)

Content		Objective Items							Subjective Items					
		Multiple Choice Questions							Very Short Answer Questions				Long Answer Questions	
Area	Weight (%)	Kn (10%)	Cp (15%)	Ap (15%)	An (30%)	Sy (10%)	Ev (20%)	Total (100%)	Kn (10%)	Cp (15%)	Ap (15%)	Sk (60%)	Total (100%)	D (100%)
Module 1	20	4	6	6	12	4	8	40	2	3	3	12	20	4
Module 2	20	4	6	6	12	4	8	40	2	3	3	12	20	4
Module 3	25	5	7	8	15	5	10	50	2	4	3	16	25	5
Module 4	25	5	8	7	15	5	10	50	3	3	4	15	25	5
Module 5	10	2	3	3	6	2	4	20	1	2	2	5	10	2
Total	100	20	30	30	60	20	40	200	10	15	15	60	100	20

Kn= Knowledge, Cp = Comprehension, Ap = Application, An = Analysis, Sy = Synthesis, Ev = Evaluation, Sk = Skill, D = All Objectives Defined

Table 1 shows that a blue print essentially containing the number of questions allocated to each topic (module), complexity of items according to Bloom's et al (1956) Taxonomy, and types of items. A test blueprint should include the assessment title; a statement of the fundamental purpose for the assessment; the learning content or learning experiences of the learners to be covered by the instrument; a brief description of the clientele or population of the learners to be assessed like their educational level, programme and course, assumed knowledge or skill; the range of appropriate assessment type to be used both in terms of the formats within the proposed assessment and any other assessment practices which might run alongside the main scheduled assessment itself, such as oral assessments, interviews, practical work, among others; the intended uses that will be eventually made of the assessment scores to be discussed in line with the eventual results obtained by administration of the assessment instrument; the time and other relevant conditions that will be relevant for testing; and a detailed matrix which shows how the assessment instrument is developed (Adapted from Kolen & Brennan, 2019).

Criteria for Establishing Technical Quality of Assessment Items

Assessment is an integral component of a coherent educational experience. Therefore, the ideas and strategies in assessing student learning resources support three interrelated objectives for quality in student assessment in higher education. These three objectives are

- assessment that guides and encourages effective approaches to learning;
- assessment that validly and reliably measures expected learning: - outcomes, in particular the higher-order learning that characterizes higher education; and
- assessment and grading that defines and protects academic standards. (Crocker & Algina, 2020)

In order to guarantee that items pooled and stored in the item bank meet the objectives for higher education assessment identified (Crocker & Algina, 2020) above, the following standard criteria emphasized as standard (Brookhart, 2017) for establishing technical quality of test items are to be employed in test development. These are

1. **Cognitive Complexity:** This standard connotes the various levels of learning that can be assessed. *The assessment question will focus on appropriate intellectual activity ranging from simple recall of facts to problem solving, critical thinking, and reasoning.* The Bloom's Taxonomy of Educational Objectives (Bloom et al., 1956), list six levels of intellectual cognition (knowledge, comprehension, application, analysis, synthesis and evaluation) which assist in categorizing test questions in line with their cognitive complexities.
2. **Content Quality:** This standard is concerned with the quality of the assessment items with respect to skills, coverage and formats to be used. *The assessment questions will permit students to demonstrate their knowledge of challenging and important subject matter.*
3. **Meaningfulness:** The standard is concerned with the reality of the assessment items in relation to learners' expectations based on the course design. Attempt ought to be made to accommodate both the bright and dull learners by giving the learners insight on the items to be included in the assessment. *The assessment questions will be worth students' time and students will recognize and understand their value.*
4. **Language Appropriateness:** *The language demands will be clear and appropriate to the assessment tasks and to students.* This implies that the language used in asking questions should be simple, clear and free of any form of biases that will hinder the learners from giving their best performance in the assessment.

5. **Transformation and Generalizability:** *Successful performance on the assessment task will allow valid generalizations and achievement to be made.* Educators can make valid generalizations about achievement more easily using authentic and performance assessments such as presentations, scenarios, projects and portfolios to add dimensions to assessment rather than depending solely on traditional methods of assessments all the time.
6. **Fairness:** *Learner's performance will be measured in a way that does not give advantage to factors irrelevant to school learning; scoring schemes will be similarly equitable.* To achieve this standard, assessment questions should reflect the objectives of the unit; expectations should be clearly formulated tasks; one item should not aid in answering another; ample time for test completion should be allowed; and assignment of points should be determined before the test is administered. It is also desirable to provide constructive feedback to the learners.
7. **Reliability:** *Answers to assessment questions will be consistently trusted to represent what students know.* To ensure reliability of the assessment, items should be constructed by obeying the necessary rules that guide the particular type of items. It is also necessary to focus on the real purpose of the assessment and to give clear direction on what is required to the learners.

Item Writing

Once a test blueprint or table of specifications consisting of a matrix, or chart, representing the number of items intended for the test within each topic and level of objective is created, the item writing can commence. Quality test items must follow valid item-writing rules. Hence, there are guidelines (rules) that must necessarily be observed in writing items to be housed in the item bank in order to enhance their validity, reliability and usability. Daramola and Fasanmi (2022); and many other authors have painstakingly organized and provided these guidelines for educators use in writing quality items. Some of these rules and suggestions focused on: writing multiple choice items; writing true-false items; writing matching items; writing completion items; writing essay items and process in grading essay items. Item writers are encouraged to familiarize themselves with these rules before commencing item writing process.

Others are suggestions for writing problem solving items; authentic assessments such as performance test, open response questions, short investigation, portfolio; writing performance test and steps in rubric development. Frey, Petersen, Edwards, Pedrotti & Peyton (as cited in Ifedili & Ofoegbu, 2020) identified and compiled forty (40) different most commonly found item writing rules according to the relative importance of each rule, as measured by frequency with which measurement experts choose to advocate a rule and ranked by the number of sources presenting each rule. Whereas. Bloom et al. (1956) described in detail the intellectual outcomes of education and presented sample objectives and test items for each area. Educators especially item writers are encouraged to study, familiarize and equip themselves with these guidelines/rules as well as the detail intellectual outcomes objectives before commencing on item writing process. Item writing is closely followed by item review process.

Item Review Process

The items resulting from item writing stage commonly called the draft items go through several technical review and editing procedures to verify their technical content, to ensure that the draft items read well and do not provide clues to the correct answers (DeMars, 2018) and to ensure that the draft items are current, correct, relevant and will perform well as assessment items. Items that meet the expected quality standard are pooled, coded and housed in the item bank.

Paneling or Moderating Drafted Items

Paneling is a process of presenting the drafted items to a panel of reviewers or more than one critic while moderating is the actual process of the meeting to moderate the more extreme efforts of the item writer by exposing the work to other views (Wilson, 2016). The items to be moderated have to be in good condition as much as possible before presenting them to the panel members. The items are to be word processed, numbered and ordered, complete and lay out in standard form before the moderation process. No indication is to be given to the panelists of the authors suggested right answers to the items. But the items writers are to note the correct key while writing the items as they are to be provided readily for use when needed at the panel meeting. The number and choice of panelists depend on resources but they are to consist of subject and item development experts. The items writer whose work is being reviewed acts as a secretary to the panel while the drafted items are being discussed. The panelist present and defend their point of view while the chairperson has to be very firm in ensuring that the discussion is within the intended focus of the exercise. Confidentiality must be ensured, before, during and after the sessions. It is necessary to focus the discussion on critiquing of the work (items) rather than on the person of the item writer. Where possible, tape-recording of the session and keeping written records are to be done to enhance recording of essential points in the intellectual argument that are quite valuable during the later editing of the items.

Editing or Vetting

The panel meeting records are used by the items writers to edit or vet the draft items as scrupulously and gradually as possible bearing in mind that panelists are sometimes wrong. The vetting process according to Popham (2017) involves: cutting or adding to the stimulus material where the panel feels this is necessary; identifying words in the stimulus material that needs further explanation necessary for understanding the passage by the candidates; deleting whole items which the panelists think are trivial or too difficult; writing whole new items suggested as replacements or additions by the panelists; replacing the less successful draft distracters with ones the panel feels are plausible; writing new distracters suggested by the panel; checking the language at every turn for accuracy, consistency and clarity; and establishing a final order for the items in the set or the whole assessment items. These activities are necessary and they help in ensuring that essentially only the best distracters, keys, stems and stimulus materials emerged for trial testing for multiple choice items format especially. Also, the subjective items and their intended marking guides (rubrics where applicable) are moderated by scrutinizing them alongside the grading suggested by the item's writer by the panelists. Slightly variant versions of the same items can be tried out on different groups where the need arises.

Field or Trial Testing

Trial testing of items is usually carried out on population of same standard of learners on whom the assessment is designed for. These are the ones who are at the same level of study as those for whom the assessment items have been written. Popham (2017) opined that field testing is indispensable for four reasons namely: 'debugging' the whole assessment, to get rid of errors; as a check on assessment-length and timing; and as a check on the vocabulary level used in the assessment stimulus and item material. The size of the trial population is usually the largest possible population given the available resources. Also, the population for the trial testing is to be close in make-up to the one for which the assessment is ultimately intended in terms of age; level of schooling; gender composition; experience of the subject matter which forms the basis of the

assessment. To ensure security and prevent leakage of the items, only people connected with or hired by the project team will handle papers and supervise trial sessions. This is because the more people who see the items, the more likely copies are to disappear and leak. Ample time is given to students to work through and complete the assessment items in the field trial session. This helps in ensuring that the candidates react maximally to the items. The result of the trial testing guides in setting parameters in item selection and timing during the actual assessment process. Whenever field trial testing is not possible, the test developer should be especially-thorough in checking the items based on the items review process outcomes before they are housed in the item bank. In this case, more professionals, both subject matter and test experts, are to look at the draft forms of the assessment items and offer critical comments that guide further moderation of the items.

Item Analysis

This consists of statistics which usually show the facility of each item for multiple-choice items. It basically shows the percentage of the whole test-taking population which get it right; the discrimination index of each item - how well the keyed answer distinguishes between students of high ability and those less able; the response level for each item - how many actually attempted it, right or wrong; the criterion score on each item - the mean score of all those who did attempt it; and whether any distracter did not function well - attracted too few candidates, or a predominance of those of high ability (Brookhart, 2017). Crocker & Algina (2020) noted that at least five other things are under analysis during this vital stage of the whole development process. These are that information is obtained about: the parts of an item, especially in multiple - choice format with its stems, keyed answers and distracters; the integrity or usefulness of the item as a whole; the performance of the item as a discrete assessment element; the performance of an item with regard to other items in the same set or assessment; the integrity or usefulness of the assessment as a whole. The wide range of choices to be made and all the adjustments in the item analysis is still guided by the curriculum and the objectives to be assessed which are already embedded in the table of specification or blueprint. That is, the blueprint is useful and master guide in the entire process of development of items for quality banking.

Final Editing

This second stage of editing is done using the analysis results to scrutinize each item concerning rejection, retention or modification. The modification of items is limited to minor changes to distracters only for multiple-choice items. Any item that calls for major change is to be stepped down for further modification and subsequent trial testing. The clean items are pooled and banked in the item bank ready for use as at when needed. The preparation of items for use is greatly facilitated if the items are properly recorded. The recording may be based on the type of items to be banked, the learning outcomes measured, the difficulty levels of the items and the subject matter measured. It is necessary that any institution developing item bank designs a model for coding items to be housed in the item bank for easy retrieval or the whole process will amount to wasted effort. An example of a coding system developed and suggested for NOUN assessment system for the course Foundation of Education is illustrated by Iroegbu and Ugwuanyi (2019). This code facilitates the identification of items in terms of all the necessary parameters. Arhin (2017) surmised that item codes facilitate selection of items to fit a given test blueprint or area of content; helps to get the desired balance between different skill levels; ensure that items have the graphics or other stimuli they need; and that one item does not "give away" the answer to one or more other items. The code may include information such as institutional level, subject area, instructional objective, content topic, and item statistics such as difficulty and discrimination

indexes (Boone & Staver, 2017). The coding makes it possible to select items to build an assessment exercise that matches a particular set of assessment specifications for use either for continuous assessment purposes or for summative assessment. The coded information concerning each item also aids in arranging the items in the assessment such as by objective, order of difficulty, level of study, programme, course and other desired arrangements.

Item Banking

All potential entries into the item bank must meet technical quality, match the curriculum and be free of potential bias. This process involves coordination between the test development experts and subject matter specialists for the various courses and programmes. The item bank should provide a scale of relative difficulty of tasks that covers multiple grade levels and skills within content areas. Also, information on relative difficulty of different assessment tasks within and across levels of study is to be provided. Other relevant parameters that are necessarily included are course (subject) and the topic within that course, degree of difficulty, cognitive level, information on previous usage if any and similarities are to be indicated. Items are coded based on predetermine nomenclature. Once all these processes are concluded, the items are ready for banking. Item banking takes a number of steps and strategies. Some of these steps and strategies according to Arhin (2017) are briefly presented below:

- **Item Entry and Storage:** This can be done by keying in on a computer keyboard; by reading a disk or tape already created using a word processing programme; from hard copy by use of document reader or the item banking programme may provide algorithms that create items.
- **Graphics and Special Characters:** The handling of graphics is very important in some testing programmes such as mathematics, science, and occupational skills. Many programmes do not have the capability of handling graphics within the item file rather they are in separate file either computerized or in hard copy. Moreover, some programmes are limited as to the extent to which they will handle special characters such as mathematical symbols, foreign alphabets, accents, and so forth. Nevertheless, graphics and special characters are to be handled during item entry by using appropriate programmes.
- **Test Limitations:** Many programmes limit the number of items that may be stored in a given bank. Some limits as to how many lines of text are permitted in either the stimulus or option, whereas others limit in both.
- **Size of Item Bank:** While some programmes are designed to handle a limited number of items in a disk, others use multiple disks or auxiliary storage unit.
- **Types of Items Used:** Some programmes are designed to handle multiple test formats while some others handle single test format. The adopted programme must handle the types of items to be used for the assessment by the institution.
- **Ease of Editing or Adding New Items to the Bank:** This is of great importance to enable effective changes in the item bank as at when necessary. This can be done following same considerations as for item entry.
- **Classification/coding:** In order that the item can be retrieved to create test, there must be a capability to retrieve the items as necessary. To achieve this, items are coded in some fashion by providing certain information. The codification of items are done before storage in the item bank (Mohammed & Kabir, 2025), and for ease of computerization and retrieval. To enable assembling of items to fit the desired coverage of the assessment. The item code often indicates content, skill level and assessment type. Each item usually must have a unique identification number that may be the accession number within the classification scheme.

Other relevant information such as author, reviewers, date of revision, and references are also provided in item bank.

Conclusion

Item bank is a planned library of assessment items pooled through cooperative efforts under the auspices of an institution for use in the assessment of students learning outcomes in partial fulfillment of the requirements of the teaching learning process. It is designed to fulfill certain predetermined purposes notably to enrich the instructional process and to judge learners in terms of instructional efforts. This is necessary because most tutors especially in open and distance learning settings are supposed to prepare a large number of quality items in different topics of prescribed courses in different programmes for various assessment purposes such as Tutor Marked Assignments (TMAs) (continuous assessment) and End of Semester Examination (ESE) (summative assessment) items. But many tutors saddled with such responsibility lack adequate know-how coupled with inadequate time at their disposal to effectively carry out this responsibility. Building item bank is a gradual process that progress over time and its usefulness becomes obvious when it becomes fully functional.

In order to ensure quality of questions papers and for effective and responsive assessment of learning outcomes, development of quality items pools in different courses and programmes has to be entrusted to experience tutors who are trained with the content and technique of framing items. These items when pooled and deposited in item bank; aids immensely in the assessment of students learning outcomes and will go a long way in transforming assessment processes in open and distance learning in Nigeria. However, it should be noted that this transformation can only be achieved if the items are developed according to the assessment blueprint pre-determined framework based on content, objectives, item writing rules and other relevant parameters specified for the assessment purposes. The development and use of test blueprint in item construction can be seen as one way of enhancing the quality of items which leads to improved overall quality of assessment. It ensures that the emerged items will still be able to satisfy the demands of the assessment purposes. Coding of items is necessary and is done to facilitate their use or the item bank will lack utility. Coding makes it possible to select items in accordance with any table of specifications in the particular area covered by the item bank. Item bank no doubt is a viable asset in the deployment of meaningful assessment many institutions especially in open and distance learning settings characterized by large students.

Suggestions

In order to develop a robust, high-quality, sustainable item bank aligned with global best practices in educational measurement, the following suggestions are made.

1. Establish clear psychometric standards by aligning item development with Classical Test Theory or Item Response Theory to ensure reliability, validity, and fairness across diverse student populations.
2. Integrate Bloom's Taxonomy to ensure items comprehensively assess cognitive domains, from knowledge recall to higher-order thinking like analysis and evaluation.
3. Adopt a systematic item review process, involving subject matter experts, test developers, and data analysts to identify biases, ambiguities, and curriculum misalignments.
4. Utilize technology and digital platforms for item bank construction, storage, tagging, calibration, and adaptive testing to enhance accessibility, security, and efficiency.
5. Incorporate continuous validation cycles, using statistical item analysis (e.g., difficulty index, discrimination index) from pilot testing to refine and update the item bank regularly.

References

- Adebayo, O. O., & Adetunji, A. M. (2020). Building a quality item bank for national examinations: A framework for Nigerian educational assessment bodies. *Journal of Educational Assessment in Africa*, 15(2), 45–60.
- AERA, APA, & NCME. (2018). *Standards for educational and psychological testing*. American Educational Research Association.
- Ajayi, K. O., & Lawal, A. M. (2019). Psychometric properties of items in teacher-developed tests: Implications for item bank development. *Nigerian Journal of Educational Research and Evaluation*, 18(1), 12–24.
- Alade, O. M., & Ojo, T. S. (2021). Development and validation of an item bank for secondary school mathematics. *Journal of Educational Measurement and Evaluation*, 6(2), 67–81.
- Ali, A. M., & Musa, I. A. (2018). Building quality item banks for large-scale assessment: Challenges and prospects. *International Journal of Educational Assessment and Evaluation*, 10(3), 55–70.
- Arhin, A. K. (2017). Reliability and validity of test items in a high-stakes environment: A precursor for item banking. *Educational Measurement and Testing*, 13(2), 45–60.
- Bature, I. J., & Sunday, C. E. (2020). A systematic approach to item bank development for e-assessment in secondary schools. *Contemporary Educational Technology*, 11(4), 345–359. <https://doi.org/10.30935/cet.668923>.
- Bello, R. S., & Yusuf, A. (2019). Item banking and computerized adaptive testing: A pathway for improving assessment in Nigerian schools. *Journal of ICT in Education*, 4(1), 80–95.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals (Vol. Handbook I: Cognitive Domain)*. David McKay Company.
- Boone, W. J., & Staver, J. R. (2017). *Advances in item response theory: Research and applications*. Springer. <https://doi.org/10.1007/978-3-319-67280-0>.
- Brookhart, S. M. (2017). *How to create and use rubrics for formative assessment and grading*. ASCD.
- Brown, G. T. L., & Harris, L. R. (2021). Chapter 6: Item writing and design: Improving assessment quality. In G. Brown & L. Harris (Eds.), *Handbook of human and social conditions in assessment*. Routledge. <https://doi.org/10.4324/9781315694381-6>.
- Camilli, G., Shepard, L. A., & Popham, W. J. (Eds.). (2017). Downing, S. M., & Haladyna, T. M. (Eds.). (2016). *Handbook of test development (2nd edition)*. Routledge. <https://doi.org/10.4324/9781315774121>.
- Chikwere, C. A., & Opara, J. A. (2023). Building and managing an item bank for science achievement tests: A Nigerian perspective. *Journal of Assessment and Evaluation in Education*, 9(1), 88–102.
- Crocker, L., & Algina, J. (2020). *Introduction to classical and modern test theory*. Cengage Learning.
- Daramola, M. A., & Fasanmi, S. O. (2022). Item analysis as a tool for quality assurance in test construction and item banking. *West African Journal of Educational Research*, 25(1), 123–137.
- DeMars, C. E. (2018). *Classical test theory*. Routledge. doi.org/10.4324/9781315161269.
- Downing, S. M., & Haladyna, T. M. (Eds.). (2016). *Handbook of test development (2nd edition)*. Routledge. <https://doi.org/10.4324/9781315774121>.

- Eze, P. N., & Mbanusi, C. O. (2018). Quality control in item bank development: Lessons from tertiary institutions in Nigeria. *International Journal of Educational Research and Development*, 8(4), 58–71.
- Ifedili, C. J., & Ofoegbu, F. I. (2020). The role of item response theory in the development of a standardized item bank. *Nigerian Journal of Educational Foundations*, 22(2), 110–124.
- Iroegbu, V. I., & Ugwuanyi, C. S. (2019). Challenges in the development and management of item banks in Nigeria. *International Journal of Educational Studies*, 3(2), 66–78.
- Isaacs, T., & Gultig, J. (Eds.). (2021). *Assessment in education: Principles, policies, and practices in the 21st century*. Pearson South Africa.
- Kamba, M. A., & Okonkwo, A. P. (2023). Digital item banking for formative and summative assessments in secondary education. *Journal of Educational Technology and Assessment*, 7(1), 33–49.
- Kolen, M. J., & Brennan, R. L. (2019). *Test equating, scaling, and linking: Methods and practices (3rd edition)*. Springer. <https://doi.org/10.1007/978-3-030-21847-1>.
- Liu, Q., & Yin, Y. (2019). Using large-scale assessment data to inform instruction and assessment item bank development. In L. M. English (Ed.), *Assessing higher order thinking in mathematics and science*. Springer.
- Madaus, G. F., Russell, M. K., & Higgins, J. (2018). *The paradoxes of high-stakes testing: How they affect students, their parents, teachers, principals, schools, and society*. IAP Publishers.
- McMillan, J. H. (2018). *Classroom assessment: Principles and practice for effective standards-based instruction (7th edition)*. Pearson.
- Mohammed, A. A., & Kabir, A. A. (2025). Building robust item banks for adaptive testing in West African secondary schools. *African Journal of Educational Research and Development*, 12(1), 14–29.
- Nwankwo, A. A., & Uzoechina, G. O. (2017). The psychometric evaluation of test items as basis for constructing item banks. *Journal of Educational Research and Policy Studies*, 11(2), 87–101.
- Popham, W. J. (2017). *Classroom assessment: What teachers need to know (8th edition)*. Pearson.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15(1), 72–101.
- Tao, J., & Chang, H. H. (2020). Chapter 4: Item bank development and management for computerized adaptive testing. In S. Jiao & W. J. Boone (Eds.), *Modern measurement: Theory, principles, and applications*. Springer.
- Usman, A. A., & Musa, J. (2019). A framework for constructing sustainable item banks in teacher education programs. *Nigerian Journal of Teacher Education and Evaluation*, 7(2), 134–147.
- Wilson, M. (2016). *Constructing measures: An item response modeling approach*. Routledge.