Journal of Teaching and Teacher Education
ISSN (2210-1578)
J. Tea. Tea. Edu. 11, No. 1 (Jan-2023)

http://dx.doi.org/10.12785/jtte/110101

Enhancing Test Development and Item Writing Skills of Subject and Vocational Teachers in Government-Owned Junior Secondary Schools in Jigawa State, Nigeria

Ali Abdullahi Taura [1] and Abdul-Wahab Ibrahim [2]

[1, 2] Department of Education, Sule Lamido University Kafin Hausa, Jigawa State, Nigeria

Received 30 Oct 2022, Revised 26 Nov 2022, Accepted 5 Dec 2022, Published 01 Jan 2023

Abstract: This study conducted a capacity building training programme to determine whether there exists effect of treatment on teachers' knowledge of basic test development and construction skills in public Junior Secondary Schools in Jigawa state, Nigeria. Also, it examined if there exists effect of treatment on subject and vocational teachers' item writing and test analysis skills, teachers' ability to categorize and interpret test scores for learning improvement in public Junior Secondary Schools in the state. This is with a view to enhance teachers' test construction competence and improve both validity and equity in achievement testing. An equally randomized quasi-experimental research design was employed in the study. The population studied comprised all male and female teachers in government-owned Junior Secondary Schools in all 27 Local Government Areas (LGAs) of Jigawa state, Nigeria. 520 secondary school personnel constituting 400 subject and vocational teachers, 40 Guidance Counsellors and 80 School Administrators constituted the sample in the study. Three self-developed research instruments were used in the study to collect data. These are: "Test Development Questionnaire (TDQ); Item Writing Questionnaire (ITQ), and Impact Assessment Questionnaire (IAQ)". Independent t-test statistics was used to test hypotheses in the study at 0.05 level of significance. The results showed that there existed a significant effect of treatment on subject and vocational teachers' knowledge of basic test development and construction skills in Junior Secondary Schools in Jigawa state. Also, there was a significant effect of treatment on subject and vocational teachers' item writing skills and teachers' skills in item and test analysis in Junior Secondary Schools in the state. However, a non-significant effect of treatment on subject and vocational teachers' ability to categorize and interpret test scores for learning improvement in Junior Secondary Schools in Jigawa state existed. It was concluded that continuous improvement in test development and construction is imperative for the 21st Century classroom teachers in the country. The study recommended that there was an urgent need to strengthening teacher education in view of competence-based test and item writing skills enhancement amongst secondary school teachers in Nigeria.

Keywords: Test development, Test construction, Test analysis, Item writing, Skills

INTRODUCTION

In all educational settings, teachers have taken active part in the construction of tests without the help of professional test experts, reviewers and item writers. It is incumbent on the classroom teacher to assess the extent his students have mastered the intended learning outcomes. To do this, he uses achievement tests. Some achievement tests assess instructional objectives common to all students across all education spectrums in the country, while others measure learning objectives specifically taught by classroom teachers. This is because testing is an integral and complimentary part of the teaching and learning processes (Ibrahim, 2022). It is therefore imperative that every teacher is proficient in constructing achievement tests in his subject area.

Studies in the field of education have established the influence of testing on the process of teaching and learning (Chapman & Snyder, 2010; Wall, 2010; Wall & Alderson, 2013). Test results across the world are used as an indicator of the performance of teachers, schools, and the accountability of the education system. In many education systems around the world, the extent to which achievement test contribute to improved learning and instruction is determined largely by the principles underlying their development. Tests can direct students' attention toward the objectives of instruction or away from them. They can encourage students to focus on a

E-mail address: alia.taura@slu.edu.ng, abdulwahab.ibrahim@slu.edu.ng

limited aspect of the course content or direct their attention to all-important areas. They can reward superficial learning or require depth of understanding. They can provide dependable information for instructional decisions or they can provide biased and distorted information. If a test is to be successful, careful planning must precede its development and construction (Ibrahim, 2016; Ibrahim, 2022).

According to Dibu-Ojerinde (2012), one must consider the objectives to be measured, the purpose the scores are to serve and the conditions under which the testing will occur. The first step in building a classroom test is for the classroom instruction to specify the goals, objectives, and desired outcomes. These goals must be stated in behavioural terms, in terms that clearly specify the type of behaviour that the student is expected to exhibit. Similarly, Sidhu (2012) stated that a test has to be prepared through many stages. The first thing to know is why the test is being developed, constructed and given. For this purpose, we have to prepare educational objectives for the sub-units, units and the courses. Then, we have to enlist in detail the specific outcomes of learning which will determine the specific pieces of information, thinking, application, to mention only a few, for which items have to be written. The idea of educational objectives provides to us a new way of thinking about teaching, learning and testing. Likewise, Adediwura (2012) conceptualized educational objective as that at which the teacher aims at testing. The educational objectives state that what the teacher is trying to achieve, where he is trying to go and/or why he is doing what he does. An objective provides both the teacher and the student a direction concerning the subject content and the mental process expected of the student.

Further, Faleye (2012) asserted that educational objectives fall into one of the three main headings that are popularly referred to as domains in education. These are: cognitive objectives, which deal mainly with rational, intellectual thought processes and concerned with knowledge outcomes, information, intellectual skills and abilities; the affective objectives, which emphasize attitudes, interests, values, appreciations, feelings and emotions as well as modes of adjustment; and the psychomotor objectives, which are concerned with muscular and motor skills, the manipulation of materials or objects, and activities which require muscular coordination. Each of these three domains has content and process objectives. As a result of the great value placed on testing, some believe that testing provides incentives to students and their teachers to improve test performance. The society accessibility to test results also pushes schools to provide any support necessary for the same purpose. These efforts therefore, are believed to help raise the level of achievements.

For instance, Ibrahim (2018) conducted a quasi-experiment on how to improve assessment and evaluation skills of teachers in public primary and secondary schools in Jigawa State, Nigeria. The population comprised of teachers who were teaching in all government primary and secondary schools in the state.115 randomly selected participants comprised the sample in the study. Three self-constructed instruments were administered and used to collect data in the study. Frequency counts, percentiles, mean scores, standard deviations, independent t-test, Analysis of Covariance (ANCOVA), and Fisher's t-test statistical techniques were used to analyse the collected data in the study. The results revealed that there was low level of compliance to and implementation of the objectives of Continuous Assessment in the state. Also, there was a significant effect of the treatment on the teachers' skills in evaluation techniques of cognitive, affective and psychomotor domains of learning in public schools. He concluded that there was a need to train and retrain all public school teachers in the state on the concept and philosophy of Continuous Assessment in educational system in Nigeria.

In the same vein, Hamafyelto et al. (2015) studied competence of commerce teachers' test construction in public and privately-owned Senior Secondary Schools in Borno State, Nigeria. Also, the researchers investigated the relationship between commerce teachers' test construction and content validity of commerce examination questions in Senior Secondary Schools in the State. The population was 75 commerce teachers, out of which 16 of the teachers were teaching commerce as a subject in public schools, while the remaining 59 commerce teachers were teaching in private secondary in the state. Hence, no sampling technique was used to select commerce teachers as all of them were used as participants in the study. A researcher-constructed 42-item "Teachers Competence Questionnaire (TECOM-Q)" was administered to collect data from participants. Frequency counts, percentages, mean, standard deviations and Chi-Square statistics were used to analyse the data. The result of the analysis showed that there were significant relationships between teachers of commerce competence and content validity, the areas of teachers' competence in constructing examination questions was low. It was found that teachers concentrated on the lower levels of the cognitive domain (remembering, understanding applying). They concluded that workshops and seminars are needed to improve teachers' competence in test construction.

Likewise, Kazuko (2010) investigated Japanese high school mathematics teachers' competence in real world problem solving. It was revealed that participant-teachers possessed mathematical knowledge skill in written mathematical modeling tests. The researcher concluded that the quality of test questions depended on the quality of the teacher as teachers' level of competence is one of the factors that directly affects the quality of test items. It was argued that the quality of test items constructed by

teachers is closely linked with their ability to provide relevant information as regards students' performance. A well-written test helps teachers to correctly measure students' mastery of the subject matter and content taught in the classroom. On the contrary, poorly designed test items could lead to unintended learning outcomes both in terms of students' performance as well as instructional effectiveness.

Subsequent research by Amani et al. (2021) examined secondary school teachers' competences, awareness of skills and procedures for constructing quality classroom tests in Tanzania. 246 secondary school teachers who were drawn from four regions in Tanzania participated in the study. A semi-structured questionnaire was used as the data collection tool for the study. Frequencies, percentages, and charts were used to analyse the data. The findings showed that the majority of the participant teachers lacked competences for preparing quality classroom tests, particularly on the use of Table of Specification and test-item analysis. Also, teachers lacked professional support on how to prepare matching items, short answers, and multiple-choice test items. Again, a preponderant majority of teachers had never received in-service training on the subjects of assessment and testing. The researchers concluded strengthening initial teacher education in view of competence-based assessment.

In a related study, Quansah et al. (2019) investigated teachers' test construction skills in Senior High Schools in the Cape Coast Metropolis of Ghana. Samples of end-of-term examinations in three selected core subjects in the schools were randomly selected for the study. The results revealed that the teachers have limited skills in the construction of end-of-term examinations. Also, teachers were found lacking in competence content representativeness, relevance of the test, reliability, and fairness of the assessment tasks as evaluated. The researchers concluded that test specialists from recognised academic institutions should be invited to organise workshops for teachers on a regular basis to sharpen their skills on effective test construction practices.

Similarly, Agu et al. (2013) determined teachers' competencies in constructing classroom-based tests in public secondary schools in Onitsha Education Zone in Anambra state, Nigeria. The sample is made up of 543 male and female secondary school teachers selected using stratified random sampling technique from three local government areas that make up Onitsha Education Zone. A self-developed 30-item "Test Construction Skill Inventory (TCSI)" was used to collect data from participants. Factor analysis was used to analyse the data. The results showed that 25 items were found to be factorially valid as secondary school teachers found almost all the 25 items important skills for quality classroom-based test construction. The researchers concluded that TCSI was an important measure for determining the secondary school teachers' test construction skill in Anambra State, Nigeria.

From the above delineation, it is believed that the quality of teachers in any given educational business is sine-qua-non. Thus, teacher's quality is significantly and positively correlated with students' performance. It is therefore, expected that teachers must possess technical competence and professional skills through a well-coordinated teachers capacity building and training intervention or/and education programmes that can meet the various challenges relating to efficient learning environment. As a result of the negative consequences of current testing practices throughout the country, the subject and vocational teachers have been urged to revisit the purpose and value of assessment in the teaching and learning processes. It was noticed that testing mainly serves the purpose of summarizing students' achievement by giving grades and reporting marks to students at the end of teaching and learning. In most cases, it is not expected that teachers will have a greater deal of time available in which to build tests, but it is certainly not recommended that they "knock something out" two or three days before tests are to be used. Whether as a pre-test, weekly test, terminal test or promotion test, a classroom teacher is expected to be minimally equipped to develop test items, grade test answers and report test results, hence this capacity building study.

However, it has also been argued that testing only motivates teachers and students to work towards performance goals rather than learning goals. Accordingly, the increase in scores, especially in high-stakes testing context, most likely indicates teachers' and students' familiarity with test requirements and formats rather than the real improvement in learning. Also, that the ability of standardized testing to measure the whole range of knowledge and skills that students are supposed to acquire is questionable. Teachers, in order to drive up test scores, tend to teach to the test, focusing on what is to be tested and developing test taking strategies, but ignoring those skills not covered in the test. Hence, it is imperative that every teacher is proficient in developing, constructing, and writing test items in his subject area. It is against this background that this study investigated test development and item writing skills of subject and vocational teachers in public secondary schools in Jigawa state. By so doing, the study will provide information on suitability and ease of use of the different test development, construction and item writing methods. It will also provide information on basis for the comparison of the effectiveness of the different test development, construction and item writing methods especially on how and when to use each of the methods. Also, the outcome of the study will assist teachers to have a proper balance of psychometric and other properties of available test item formats, and hence enhance quality decision making on testing. Not only this, but also the result will assist subject and vocational teachers to optimize the difficulty level of test items, vis-a-vis selecting suitable number of distracters appropriate for Junior and Senior Secondary level students in Jigawa state and the country. Further, the

results of this study will assist in identifying areas where large-scale achievement test may be improved to enhance fairness and accuracy. The results will enhance both validity and equity in testing.

As a corollary to the above, the main objective of this study is to conduct capacity building training programme and determine test development and item writing skills of subject and vocational teachers in government-owned Junior Secondary Schools in Jigawa state. In order to achieve this goal, the specific objectives of the study are to:

1. Determine test development, construction, and item writing skills of subject and vocational teachers in junior secondary schools in Jigawa state.
2. Acquaint subject and vocational teachers on basic concepts and procedures in test development, construction, and item writing.
3. Develop subject and vocational teachers' skills in item and test analysis.
4. Categorize and interpret test scores for learning improvement.

Research Hypotheses

1. There is no significant effect of treatment on subject and vocational teachers' knowledge of basic test development and construction skills in Junior Secondary Schools in Jigawa state.
2. There is no significant effect of treatment on subject and vocational teachers' item writing skills in Junior Secondary Schools in Jigawa state.
3. There is no significant effect of treatment on subject and vocational teachers' skills in item and test analysis in Junior Secondary Schools in Jigawa state.
4. There is no significant effect of treatment on subject and vocational teachers' ability to categorize and interpret test scores for learning improvement in Junior Secondary Schools in Jigawa state.

THEORETICAL FRAMEWORK

The study was guided by the behaviourist and cognitive constructivist theories as well as social learning theory propounded by Pavlov, Skinner, Thorndike and Bandura respectively. A glance at the Behaviourist theories of learning finds learners, in our case, teachers in the experiment learning and receiving treatment or training on how to improve their test development and construction skills, which is to enable them become proficient in developing, constructing, and writing test items in their subject areas. Behaviourism as a learning theory focused on experimenting stimulus-response patterns of conditioned behaviours, reinforcement and behaviour shaping (McDevitt & Ormrod, 2013; Al-Mahdi & Al-Wadi, 2015).

Generally behaviourism is based on three assumptions common to most of the theorists (Pavlov, Thorndike, Skinner, Bandura and others). The three assumptions central to explaining the learning process hold that (1) learning is a change in behaviour; (2) the environment (stimulus) shapes behavior; and (3) the closeness in time for occurrence of event creates a firm bond. Behaviourism focuses mainly on the objectively observable aspects of learning. For behaviourists, when behaviour is reinforced, it increases the likelihood of it recurring. If the behaviour is punished it may never recur. It also encourages connections to be made and creates opportunities for associations (Tchombe, 2011).

Likewise, the cognitive constructivist as a learning theory challenged the behaviouristic atomization view of knowledge. According to constructivists, human cognition develops through two processes namely assimilation, which includes process of assimilating external actions into thoughts and fitting new mental models into the existing mental structures; and accommodation, which includes the process of structuring the adopted mental material in the mind. Accordingly, the latter process develops through four major periods of human life: (i) the sensory motor period; (ii) the pre-operational period; (iii) the concrete-operational period; and (iv) the formal operational period (McDevitt & Ormrod, 2013; Al-Mahdi & Al-Wadi, 2015).

To social learning theorists, observational learning is also vicarious learning (Bandura, 1977). Bandura calls the process of social learning, modelling and provided four conditions, namely, attention, retention, motor reproduction, and motivation. The theory's central concept is reciprocal determinism, whereby the interacting factors in learning are both cognitive and environmental, acting on the learner's behaviour (Bandura, 1977). These determine not only the learner's emotional reactions but also the learner's beliefs, expectations and behavioural manifestations. To Bandura, learning is copying, modeling, observing and imitating but with some awareness of what is involved. The observational learning requires continuous reciprocal interaction between cognitive, behavioural and environmental factors (Bandura, 1977). Observation is very important in childhood and is critical for children's stimulation. Bandura (1977) states that in observational learning, the learner learns by observing the behaviours of others. The potentials and power of observational learning was advocated by Dewey (1997). As we observe the external, we also engage in self–observation which is internal to us. Bandura (1997) refers to this self- efficacy appraisal if this is done against set standards or established goals, such as lesson objectives. Two sources of self-efficacy appraisal are common amongst students; awareness of one's actual performance and when students are influenced vicariously as they see peers reinforced in their successful task

performance. Social learning theory demands teacher-guided facilitation of students' interaction in cooperative learning.

Contextually, we can see from the reviewed theories above that scholars have variously propounded different theories of learning and how these influence teaching and learning. Since the whole essence of teaching is to bring about quality learning, learning theories can provide a guide in finding new ways for improving classroom teaching especially teaching practices as it affects teachers' professional development as whole. For instance, Schunk (2011) conceptualized learning as an "enduring change in behaviour, or in the capacity to behave in a given fashion, which results from practice or other forms of experience" (p. 2). Undoubtedly, some learning theorists will disagree on the definition of learning presented here. However, it is not the definition itself that separates a given theory from the rest. The major differences among theories lie more in interpretation than they do in definition. Every teacher should endeavour to know the importance of school learning; that there are different types of learners and learning styles. Learners have been categorized as field–dependent (concrete) learners, field-independence (abstract) learners, imagery learners (learn more through images), and verbalizers (learn more through the spoken word). In the act of learning three processes take place; acquisition of new information, transformation of knowledge acquired and evaluation of the learning process (Bruner, 2008).

METHODOLOGY

Study Area

The study was being carried out in Jigawa state. Jigawa is a state in central northern Nigeria. Its capital is Dutse. Jigawa State is one of thirty-six states that constitute Federal Republic of Nigeria. It is situated in the north-western part of the country between latitudes 11.00°N to 13.00°N and longitudes 8.00°E to 10.15°E (Wikipedia, 2012).

Research Design

The research design for this study used was an equally randomized quasi-experimental research design, to investigate test development and item writing skills of both subject and vocational teachers in public Junior Secondary Schools in Jigawa state, Nigeria. This approach was used because the study conducted a capacity building training programme to determine test development and item writing skills of subject and vocational teachers in government-owned Junior Secondary Schools in Jigawa state. According to Cohen et al. (2018), quasi-experiment is a research design in which a researcher has only partial or no control over randomly assigning participants to levels of a manipulated variable of interest. Thus, the investigators recruited a study sample that shares certain characteristics by formally stating specific inclusion and exclusion study criteria when designing this study. For example, inclusion criteria include that participants must be subject teacher of either an Arts or Science or Commercial and vocational subject as Technical, Computing, and Home Management subjects and English language speaking; while exclusion criterion was that participants must not be under the five years of teaching or employment of Jigawa State Ministry of Education as teachers Junior Secondary Schools in the state. Using the classic notation system provided by Campbell and Stanley (1963, p. 6; cited in Cohen et al., 2018), a diagram to illustrate the quasi-experimental research design as used in this is as follows:

R	0_1	X_0	0_2	C
R	0_1	X_1	0_2	E

Where R represents equal random assignment into Experimental Group (E) and Control Group (C), 0_1 and 0_2 represent Pretest and Posttest observations, X_0 means no treatment for school administrators and counsellors, who formed the Control Group. Also, X_1 treatment for subject and vocational teachers in the experiment. This means that of the pool of participants, individual 1 goes to Group 1, Individual 2 to Group 2, and so forth so that there is no systematic bias in assigning the individuals. This procedure eliminates the possibility of systematic differences among characteristics of the participants that could affect the outcomes so that any differences in outcomes can be attributed to the study's manipulated variable (or variables) of interest (Keppel & Wickens, 2003). In this case the purpose is to get subject and vocational teachers views of their test development, construction and item writing experiences thus far on the achievement tests used in their respective schools and to try to ascertain whether they consider the activities they have been involved in worthwhile and to help them to become more skillful in test development, construction, and item writing as subject and vocational teachers.

Participants

The population studied comprised all male and female teachers in government-owned Junior Secondary Schools in all 27 Local Government Areas (LGAs) of Jigawa state, Nigeria. As inclusion criterion to be eligible to participate in the study, the teachers had to be employees in the government-owned schools and must have spent not less than 5 years on the job, and show willingness to take part in the study. However, 520 secondary school personnel constituting 400 subject and vocational teachers, 40 Guidance Counsellors and 80 School Administrators (Headmasters, Deputy Headmasters, Principals and Vice Principals) from the 40 selected Junior Secondary Schools make up the accessible population in the study. Also, a total of 40 Government-owned Junior Secondary Schools from 20 of the 27 Local Government Areas of Jigawa state were selected to participate in the study through stratified random sampling. Being a quasi-experimental, the design comprised one experimental group, that is, one treatment group and one control group. The 40 schools

were considered adequate representation of the state, since the school population from where the sample was drawn is run by the same agency (Ministry of Education), thus some degree of homogeneity is assumed. Furthermore, the schools within the State were stratified along junior-senior secondary school dichotomy before simple random sampling was employed to select the participants (subject and vocational teachers). From each of the 40 schools, the researchers randomly selected 20 subject and vocational teachers (10 each from the Junior Secondary Schools), thus giving a total of 400 teachers in the training experiment. Also, to be involved in the study were 40 Guidance Counsellors and 80 School Administrators (Headmasters, Deputy Headmasters, Principals and Vice Principals) from the 40 selected Junior Secondary Schools. From the sample list obtained from Jigawa State Ministry of Education, it was established that there were five recognized Educational Zones across Jigawa State namely: Dutse, Gumel, Hadejia, Kazaure, and Ringim zones. Noteworthy, unequal numbers of participants were selected among men and women teachers because of disproportionate number of male teachers to female teachers in the state Junior Secondary Schools; therefore, there was no gender bias or gender insensitivity and inequality in the study. Specifically, no gender discrimination was intended and promoted in the study as the participants were selected as they existed in Jigawa State schools. Table 1 showed the distribution of the participants according to locations, schools and gender.

Table 1: Distribution of Study Sample by Zones, School Type and Gender

Zones	Schools	Number of Schools	Teachers		School Administrators & Counsellors	Total
			Male	Female		
Dutse	Junior Secondary	4	30	10	20	60
	Schools	4	30	10	20	60
Gumel	Junior Secondary	4	30	10	10	50
	Schools	4	30	10	10	50
Hadejia	Junior Secondary	4	30	10	10	50
	Schools	4	30	10	10	50
Kazaure	Junior Secondary	4	30	10	10	50
	Schools	4	30	10	10	50
Ringim	Junior Secondary	4	30	10	10	50
	Schools	4	30	10	10	50
Total	**-**	**40**	**300**	**100**	**120**	**520**

From Table 1, it could be observed that 300 teachers representing a preponderant majority of teachers are men (75%), while the remaining 100 (25%) are women teachers that participated in the study. This represented the demographic reality of teachers in the employment of Jigawa State government at the time of this study. Also, this implied that more male subject and vocational teachers participated in the study than their female counterparts in the study. As an outlier, more subject and vocational teachers participated in the study from the Dutse Educational Zones in the study, Perhaps, this was due to the fact that Dutse, apart from being the State capital of Jigawa state, enjoyed the status of being a cosmopolitan city where virtually nearly most ethnic tribes in the country live and sojourn with their families in Dutse. Not only this, but also it has the highest numbers of public Junior Secondary Schools in the state.

Research Instruments

Three self-developed research instruments were used in the study. These are: "Test Development Questionnaire (TDQ); Item Writing Questionnaire (ITQ), and Impact Assessment Questionnaire (IAQ)". The design of the instruments was based on information obtain from relevant literature reviewed. The instruments chosen possessed the three psychometric properties of validity, reliability and usability (Ibrahim, 2017). Each of the instruments was briefly described as follows:

Test Development Questionnaire (TDQ): This is a 30-cluster item training questionnaire covering pedagogical areas of subject and vocational areas of instruction related to Junior Secondary Schools. It consists of content analysis, test blueprints, and different objective and essay test formats used in Junior Secondary schools in the state.

Item Writing Questionnaire (ITQ): This is a 24-item questionnaire covering item writing principle and item review of all subject and vocational pedagogical contents as taught in Junior Secondary Schools. It has two scales of needs and performance. The needed scale had 4-point scale response options of Highly Needed (HN), Averagely Needed (AN), Slightly Needed (SN), and Not Needed (NN), with corresponding value of 4, 3, 2, and 1 respectively. The performance scale has 3-point response options of High Performance (HP), Average Performance (AP), No Performance (NP), with a corresponding value of 3, 2, and 1 respectively.

Impact Assessment Questionnaire (IAQ): A 30-item questionnaire, which consists of three main sections with each section comprised of 10 items each. Also, it enjoys three-point Likert Scale.

Validity of the Research Instruments

The content and validity of the three instruments were established using expert judgments. Experts in Guidance and Counselling, Psychology of Education, Tests and Measurement were able to review the items in the questionnaires in terms of relevance to the subject-matter, coverage of the content areas, appropriateness of the language usage and clarity of purpose. The experts' judgments were strictly adhered to for adequate content and face validity.

Pilot Study/Reliability of the Instruments

A pilot study was conducted using Cronbach Alpha and Split-Half reliability methods to determine the internal consistency of the instruments over time. Specifically, the "Test Development Questionnaire (TDQ) showed reliability coefficient value of 0.82; $p<0.05$; while reliability coefficient value obtained for "Item Writing Questionnaire (ITQ)" was 0.75; $p<0.05$; and the "Impact Assessment Questionnaire (IAQ)" revealed reliability coefficient value of 0.79; $p<0.05$. These reliability coefficient values were considered high enough for the study.

Treatment Intervention (Training Manuals)

Training Media: The following media were deployed to achieve our training objectives.

- Projector,
- Video Tapes,
- Flip Charts, and,
- Cardboards.

Procedures: The study was conducted in three phases:

Phase 1: Pre-treatment Assessment (Baseline Assessment):

This phase has three sub-phases:

The First Sub-Phase: Pre-test assessment was given to all teachers in both areas of pedagogy as screening test for the 520 participants, which includes 400 teachers, 80 school administrators and 40 Guidance & Counsellors to select those that were qualified for the main study. The instruments "Test Development Questionnaire (TDQ), and Item Writing Questionnaire (ITQ), were administered to obtain pre-test scores of all the participants.

The Second Sub-Phase: The researchers ranked the scores from the highest to the lowest so as to determine the cut-off point of the scores. Participants whose total score fall within the cut-off were further subjected to random sampling so as to determine which of them would participate in the treatment. Eventually, those that were randomly selected were contacted through the telephone numbers provided on the questionnaire. This was possible a week prior to the commencement of the treatment.

The Third Sub-Phase: This was the classification of participants into Experimental Group and Control Group using educational zones within the dichotomy of Junior Secondary Schools. The Treatment Groups were stratified into two main groups: Group I: teachers and Group II: School Administrators/Guidance Counsellors, while Group was the Control Group. Participants in Groups I and II were picked based on variables in the questionnaire. Therefore, participants whose scores were less than 150 could not be selected for the Experimental Group but included in the Control Group. To tot up, a total of 520 participants consisting of 400 teachers, 80 school administrators and 40 Guidance & Counsellors were collapsed into the Experimental Group, whereas, participants whose total scores were less than the 150 cut-off marks on the instrument formed the Control Group.

Phase 2: Treatment Phase

Treatment for the groups: The essence of the study is to enable subject and vocational teachers in Junior Secondary Schools refresh their skills in the area of Test Development and Construction as well as Item Writing with a view to improving the quality of test items. Also, the study will assist teachers to have a proper balance of psychometric and other properties of available test item formats, hence enhancing quality decision-making on testing. Therefore, the pedagogical contents of the treatment included but not limited to:

- Overview of Test Development and Construction;
- Test Theories (Classical Test Theory & Item Response Theory)
- Concept and Types of Teacher-Made Tests;
- Taxonomy of Test Development and Construction
- Test Construction Steps (Objectives & Essay Tests);
- Item Writing and Review;
- Administering and Marking of Tests (Objectives & Essay Tests)
- Item Analysis; and
- Validity, Reliability and Dimensionality.

Practically, the treatment (experiment) lasted for 6 months (two terms), while 1 month was used to conduct impact assessment with a view of identifying inadequacy, acceptable and test items to retain, modify, reconstruct or discard.

Appointment and Training of Research Assistants

The researchers recruited and train 20 Research Assistants who were graduates residing within the five Educational Zones in Jigawa state. These Research Assistants were trained to assist in distributing the training manuals and collecting the questionnaire as well as in arranging the venues for the treatment. They were well-trained for four hours at interval of two times a week for three weeks prior to the start of the study. Specifically, accurate explanation on the objectives and goals of the research as well as ensuring confidentiality of information provided by participants and avoiding undue interference with the participants' decisions during the course of the experiment was duly given to them. Research Assistant took two days prior to the administration of the instruments so that they would be maximally make use of what they have been trained for. They actively engaged in the conduct of the study and handsomely remunerated at the end of the field work which lasted for 6 months and an extra month for impact assessment making 7 months altogether.

Procedure for Data Collection

Before the commencement of the study, permission was obtained from the Jigawa State Ministry of Education, the Local Inspector of Education (LIE) of the LGA and from the respective school authorities. Also, informed consent was obtained from the participating teachers. Data collection was done in two phases. The first phase took place prior the training of the teachers in Test Development, Construction, and Item Writing, while the second phase was after the training (post-test data). The experiment (training programme) utilized the following training methods to deliver the contents: (i) Lectures; (ii) Demonstrations; (ii) Case Study Analyses; (iv) Provocative Discourse; (v) Individual/Group Assignments; and (vi) Course Embedded Assignments and Activities/Exercises. Afterwards, the copies of research instruments were personally administered to the participants with the aid of the Research Assistants. The participants were taken into confidence during the research, as the research was implemented for six (6) months and 1 month for impact assessment.

Ethical Clearance

Applications were made to the Jigawa Ministry of Education, Science and Technology (MOEST), Dutse, to conduct all phases of the work and ethical clearance was received for data collection, supported by research training for teams. Ethical issues were kept under review at whole team training meetings, which took place prior to each phase of data collection. At these meetings, a range of ethical issues associated with the conduct of the study were discussed in detail especially as part of training for the Research Assistants recruited for the purpose of this study.

Method of Data Analysis

The mean ($\bar{x}$) and standard deviation for the pretest and posttest assessment measures were first computed before employing independent t-test statistics to hypotheses in the study. Specifically, all hypotheses were tested at 0.05 level of significance.

RESULTS

Table 2: Descriptive Data of Teachers' Test Construction and Item Writing Skills

	Subject Teachers	Vocational Teachers	Control
Pretest			
$\bar{x}$	27.85	27.85	27.81
SD	2.58	2.77	2.11
N	200	200	120
Posttest			
$\bar{x}$	31.42	30.58	29.19
SD	2.81	2.56	3.56
N	200	200	120

Table 2 presented the mean scores ($\bar{x}$) for the three groups on the pretest, that is before the experiment or treatment. A comparison of the means ($\bar{x}$) revealed that there were no statistically significant differences between the groups prior to treatment. This confirmed that the groups were essential equivalent and had no basic test development, construction and item writing skills. Further, Table 1 showed the posttest means ($\bar{x}$) for the three groups. A comparison of the mean scores ($\bar{x}$) showed that there was a positive statistically significant effect of training on both test development and construction skill and item writing skill between subject teachers and vocational teachers. This means that training (treatment) improved teachers' prowess and competence on test development and construction skill as well as item writing skill.

Testing and Interpretation of Hypotheses

Hypothesis One: In the null form, the hypothesis stated that there is no significant effect of treatment on subject and vocational teachers' knowledge of basic test development and construction skills in Junior Secondary Schools in Jigawa state. Independent t-test statistics was used to analyse the data. The results of the analysis are presented in Table 3.

Table 3: Difference in Effect of Treatment on Subject and Vocational Teachers' Knowledge of Basic Test Development and Construction Skills

Experimental Group	N	Mean ($\bar{x}$) Scores	SD	t_{cal}	p-value
Subject Teachers	200	27.94	2.65	3.09	p<0.05
Vocational Teachers	200	26.02	3.91		

***Significant; df = 399, critical t = 1.649**

Table 3 showed that significant difference in subject and vocational teachers' knowledge of basic test development and construction skills exist due to the effect of treatment during training. Subject teachers gained more as they were more affected statistically significantly ($\bar{x}$ = 27.94; SD= 2.65) higher in their knowledge of basic test development and construction skills than the vocational teachers ($\bar{x}$ = 26.02; SD= 3.91) in Junior Secondary Schools in Jigawa state. Consequently, the research hypothesis was retained, which stated that there was a significant effect of treatment on subject and vocational teachers' knowledge of basic test development and construction skills in Junior Secondary Schools in Jigawa state.

Hypothesis Two: The null hypothesis stated that there is no significant effect of treatment on subject and vocational teachers' item writing skills in Junior Secondary Schools in Jigawa state. Independent t-test statistics was used to analyse the data. The results of the analysis are presented in Table 4.

Table 4: Difference in Effect of Treatment on Subject and Vocational Teachers' Item Writing skills

Experimental Group	**N**	**Mean ($\bar{x}$) Scores**	**SD**	**t_{cal}**	**p-value**
Subject Teachers	200	24.87	2.41	2.89	p<0.05
Vocational Teachers	200	24.33	3.67		

***Significant; df = 399, critical t = 1.649**

Table 4 showed that significant difference in subject and vocational teachers' item writing skills exist due to the effect of treatment during training. Subject teachers gained more as they were more affected statistically significantly ($\bar{x}$ = 24.87; SD= 2.41) higher in their item writing skills than the vocational teachers ($\bar{x}$ = 24.33; SD= 3.67) in Junior Secondary Schools in Jigawa state. Consequently, the research hypothesis was retained, which stated that there was a significant effect of treatment on subject and vocational teachers' item wring skills in Junior Secondary Schools in Jigawa state.

Hypothesis Three: In the null form, the hypothesis stated that there is no significant effect of treatment on subject and vocational teachers' skills in item and test analysis in Junior Secondary Schools in Jigawa state. Independent t-test statistics was used to analyse the data. The results of the analysis are presented in Table 5.

Table 5: Difference in Effect of Treatment on Subject and Vocational Teachers' Skills in Item and Test Analysis

Experimental Group	**N**	**Mean ($\bar{x}$) Scores**	**SD**	**t_{cal}**	**p-value**
Subject Teachers	200	25.62	2.02	3.82	p<0.05
Vocational Teachers	200	23.98	3.50		

***Significant; df = 399, critical t = 1.649**

Table 5 showed that significant difference in subject and vocational teachers' skills in item and test analysis exist due to the effect of treatment during training. Subject teachers gained more as they were more affected statistically significantly ($\bar{x}$ = 25.62; SD = 2.02) higher in their skills in item and test analysis than the vocational teachers ($\bar{x}$ = 23.98; SD = 3.50) in Junior Secondary Schools in Jigawa state. Consequently, the research hypothesis was retained, which stated that there was a significant effect of treatment on subject and vocational teachers' skills in item and test analysis in Junior Secondary Schools in Jigawa state.

Hypothesis Four: The null hypothesis stated that there is no significant effect of treatment on subject and vocational teachers' ability to categorize and interpret test scores for learning improvement in Junior Secondary Schools in Jigawa state. Independent t-test statistics was used to analyse the data. The results of the analysis are presented in Table 6.

Table 6: Difference in Effect of Treatment on Subject and Vocational Teachers' Ability to Categorize and Interpret Test Scores for Learning Improvement

Experimental Group	N	Mean ($\bar{x}$) Scores	SD	t_{cal}	p-value
Subject Teachers	200	26.98	2.76	0.09	p<0.05
Vocational Teachers	200	26.48	3.04		

***Significant; df = 399, critical t = 1.649**

Table 6 showed no significant difference in subject and vocational teachers' ability to categorize and interpret test scores for learning improvement due to the effect of treatment during training. Subject teachers ($\bar{x}$ = 26.98; SD = 2.76) were at par with their vocational counterparts statistically not significantly ($\bar{x}$ = 26.48; SD = 3.04) in their ability to categorize and interpret test scores for learning improvement in Junior Secondary Schools in Jigawa state. Consequently, the research hypothesis was rejected and the null hypothesis confirmed, which stated that there was no significant effect of treatment on subject and vocational teachers' ability to categorize and interpret test scores for learning improvement in Junior Secondary Schools in Jigawa state.

Discussion of Findings

The purpose of this study was to conduct capacity building training programme and determine test development and item writing skills of subject and vocational teachers in government-owned Junior Secondary Schools in Jigawa state. Thus, the results of the first hypothesis revealed that there was a significant effect of treatment on subject and vocational teachers' knowledge of basic test development and construction skills in Junior Secondary Schools in Jigawa state. Also, subject teachers gained more as they were more affected statistically significantly higher in their knowledge of basic test development and construction skills than the vocational teachers in Junior Secondary Schools in Jigawa state. These findings support earlier study by Ibrahim (2018) who found a significant effect of the treatment on the teachers' skills in evaluation techniques of cognitive, affective and psychomotor domains of learning in public schools. Likewise, Kazuko (2010) revealed that participant-teachers possessed mathematical knowledge skill in written mathematical modeling tests. Similarly, Hamafyelto et al. (2015) reported a significant relationship between competence of commerce teachers' test construction in public and privately-owned Senior Secondary Schools in Borno State, Nigeria. There were significant relationships between teachers of commerce competence and content validity; the areas of teachers' competence in constructing examination questions were low. It was found that teachers concentrated on the lower levels of the cognitive domain (remembering, understanding, and applying. Whereas, the findings of this study corroborated Quansah et al. (2019) investigation on teachers' test construction skills in Senior High Schools in the Cape Coast Metropolis of Ghana, where it was revealed that the teachers have limited skills in the construction of end-of-term examinations, while teachers were found lacking in competence content representativeness, relevance of the test, reliability, and fairness of the assessment tasks as evaluated. It is plausibly explained that the knowledge of test development and construction, acquired by both subject and vocational teachers during training enhance their capacity and broaden their knowledge of test development and construction, hence this finding.

The results of second hypothesis showed that there was a significant effect of treatment on subject and vocational teachers' item wring skills in Junior Secondary Schools in Jigawa state. Further, it was found that subject teachers gained more as they were more affected statistically significantly higher in their item writing skills than the vocational teachers in Junior Secondary Schools in Jigawa state. These findings are in consonant with Ololube (2018) and Onyechere (2018) who reported that most teachers construct poor items which actually failed to function as it was supposed to. Some teachers, acknowledging that they have weak test construction skills resort to past or already existing questions to assess students (Onyechere, 2018). Similar finding was reported by Amedahe (2019), who found that Senior High School teachers in the Central Region of Ghana have inadequate skills in constructing both essay and objective type tests. Every classroom teacher is expected to possess and apply requisite skills in construction of good items for class assessments. A good test item must be both valid and reliable. A test is valid if it is suitable for the intended purpose. On the other hand, a test is reliable if it measures what it is supposed to measure consistently under all conditions (Ibrahim, 2016). Teachers today, perhaps more than ever before, have a need to be knowledgeable consumers of test information, constructors of assessments and protocols, and even teachers about testing (Rudner & Schafer, 2012).

The results of third hypothesis indicated that there was a significant effect of treatment on subject and vocational teachers' skills in item and test analysis in Junior Secondary Schools in Jigawa state. Further, it was revealed that subject teachers gained more as they were more affected statistically significantly higher in their skills in item and test analysis than the vocational teachers in Junior Secondary Schools in Jigawa state. These findings are not surprising as they are consistent with Amani et al. (2021) who discovered that the majority of the participant teachers lacked competences for preparing quality classroom tests, particularly on the use of Table of

Specifications and test-item analysis. They found a significant relationship between secondary school teachers' competences, awareness of skills and procedures for constructing quality classroom tests in Tanzania.

The results of fourth hypothesis showed that there was no significant effect of treatment on subject and vocational teachers' ability to categorize and interpret test scores for learning improvement in Junior Secondary Schools in Jigawa state. Further, it was revealed that subject teachers were at par with their vocational counterparts statistically not significantly in their ability to categorize and interpret test scores for learning improvement in Junior Secondary Schools in Jigawa state. These findings are in line with Quansah et al. (2019) who reported that teachers were found lacking in competence content representativeness, relevance of the test, reliability, and fairness of the assessment tasks as evaluated. In similar vein, Agu et al. (2013) showed that 25 items were found to be factorially valid as secondary school teachers found almost all the 25 items important skills for quality classroom-based test construction.

CONCLUSION AND RECOMMENDATIONS

On the basis of the findings of this study, it can be concluded that there existed a significant effect of treatment on subject and vocational teachers' knowledge of basic test development and construction skills in Junior Secondary Schools in Jigawa state. Also, there was a significant effect of treatment on subject and vocational teachers' item writing and test analysis skills in Junior Secondary Schools in Jigawa state. However, a non-significant effect of treatment on subject and vocational teachers' ability to categorize and interpret test scores for learning improvement in Junior Secondary Schools in Jigawa state existed. Therefore, the following recommendations were made namely continuous improvement in both our measurement methods and test development and construction is imperative for the 21st Century classroom teachers in the country. Further, there is an urgent need to strengthening teacher education in view of competence-based test and item writing skills enhancement amongst secondary school teachers. Also, test specialists from universities in the country should be invited to organise training workshops for teachers on a regular basis to sharpen their skills on effective test construction practices.

ACKNOWLEDGEMENTS

The researchers expressed their deepest gratitude to the Chairman, University Research Committee (URC), the Vice Chancellor, Professor Lawal Sani Taura (NCE, B.Sc., M.Sc., PhD., MNAMP), Sule Lamido University Kafin Hausa, Jigawa State, Nigeria, for recommending our proposal for this study for approval of Tertiary Education Trust Fund (TETfund). Also, we expressed our appreciation to TETfund for making available the Research Grant to carry out this study. Not only this, but also, we express our effusive gratitude to the Jigawa State Ministry of Education, Science and Technology (MOEST), as well as all teachers, school administrators, Guidance and Counsellors who out of their limited times, willingly find it worthwhile to participate in this study. We say a big thank you to you all!

REFERENCES

Adediwura, A. A. (2012). Development of standardized achievement tests. In E.R.I. & O.O. Dibu-Ojerinde (Eds.), *Educational test & measurement* (pp. 191-210). Obafemi Awolowo University Press.

Afolabi, E.R.I. (2012). Test score contaminants. In E.R.I. & O.O. Dibu-Ojerinde (Eds.), *Educational test & measurement* (pp. 172-191). Obafemi Awolowo University Press.

Agu, N.N., Onyekuba, C., & Anyichie, A.C. (2013). Measuring teachers' competencies in constructing classroom-based tests in Nigerian secondary schools: Need for a test construction skill inventory. Educational Research and Reviews, *8*(8), 431-439.

Ali, A.A. (2019). *Basic research skills in education*. Orient Printing and Publishing.

Al-Mahdi, O., & Al-Wadi, H. (2015). Towards a sociocultural approach on teachers' professional development in Bahrain. *Journal of Teaching and Teacher Education*,*3*(1), 89-100.

Amani, J., Kitta, S., Kapinga, O.S., & Mbiliny, C. (2021). Secondary school teachers' knowledge on procedures for constructing quality classroom tests in Tanzania. *Universitepark Bulten, 10*(1): 40-54. www.unibulletin.com

Anderson, M., &Arsenault, J. (1998). *Making the writing and research connection with the I-search process: A how-to-do-it manual for teachers and school librarians*. Neal-Schuman.

Bandura, A. (1997). *Social learning theory*. General Learning Press.

Bruner, J.S. (2008). *Beyond the information given*. Allen and Unwin.

Campbell, D., & Stanley, J. (1963). Experimental and quasi-experimental designs for research. In N. L. Gage (Eds.), *Handbook of research on teaching* (pp. 1-76). Rand McNally.

Chapman, D. W., & Snyder, C. W. (2010). Can high stakes national testing improve instruction: Re-examining conventional wisdom. *International Journal of Educational Development, 20*(2), 457-474.

Cohen, L., Manion, L., & Morrison, K. (2018). *Research Methods in Education*. Routledge.

Dewey, J. (1997). Experience and education. Touchstone.

Dibu-Ojerinde, O.O. (2012). Basic concepts: Testing, measuring, assessment and evaluation. In E.R.I. & O.O. Dibu-Ojerinde (Eds.), *Educational test & measurement* (pp. 1-20). Obafemi Awolowo University Press.

Ebinye, P.O. (2011). Problems of testing under the continuous assessment programme. *Journal of Qualitative Education*, *4*(1), 12-19.

Esomonu, N.P. (2012). Assessment of non-cognitive behaviours in secondary schools for national development. In R. C. Ebenebe & L.R. Akudolu (*Eds*.). Education for national development and integration (pp. 83-93). John Best Publishing.

Faleye, B. A. (2012). Uses of tests. In E.R.I. & O.O. Dibu-Ojerinde (Eds.), *Educational test & measurement* (pp. 21-30). Obafemi Awolowo University Press.

Hamafyelto, R.S., Hamman-Tukur, A., & Hamafyelto, S.S (2015). Assessing teacher competence in test construction and content validity of teacher made examination questions in commerce in Borno State, Nigeria. *Education*, *5*(5), 123-128. DOI: 10.5923/j.edu.20150505.01

Ibrahim, A., Shogbesan, Y.O., & Iliyasu, A. (2022). Testwiseness, digital accessibility, and attitudinal disposition to computer-assisted testing among undergraduate students in north-west, Nigeria. *Journal of Digital Learning and Education*, *2*(1), 1-12. DOI: 10.52562/jdle. v2i1.324.

Ibrahim, A. (2018). Improving assessment and evaluation skills of public-school teachers in Jigawa State, Nigeria. *Journal of Education and Practice*, *9*(11), 22-32. www.iiste.org

Ibrahim, A. (2016). Measuring differential frequency of option response patterns in four-five options multiple-choice test item among undergraduate students in Nigeria. *IOSR Journal of Research & Method in Education (IOSR-JRME)*, *6*(2), 42-48.

Kazuko, W. (2018). Assessing teachers' testing skills in Elt and enhancing their professional development through distance learning on the net. *Turkish Online Journal of Distance Education*, *5*(1), 113-125.

Kazuko, J.W. (2010). *Japanese High School Mathematics Teachers Competence in Real World Problem Solving*. Keto Academy of New York and Teachers College Columbia University.

Keppel, G., & Wickens, T. D. (2003). Design and analysis: A researcher's handbook. Prentice Hall.

Leighton, J. P., & Gokiert, R. J. (2015). *The cognitive effect of test item features*. http://www.education.ualberta.ca/educ/psych/crame/files/NCMEfinal

McDevitt, T. M., & Ormrod, J.E. (2013). *Child development and education*. Pearson International Ltd.

Onyechere, I. (2018). New face of examination malpractice among Nigerian youths. *The Guardian Newspaper* online. www.theguardiannigeria

Paulson, T.U. (2013). *The study of teaching*. Holt, Rinehart and Winston Inc.

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.

Rudner, L.M., & Schafer, W.D. (2002). *What teachers need to know about assessment*? National Education Association.

Sidhu, K.S. (2012). *New approaches to measurement and evaluation*. Sterling Publishers Private Limited.

Silker, R.T. (2003). *Teachers and tests*. Basil Blackwell.

Schunk, D. (2011). *Approaches to training and development*. Addison-Wesley.

Tchombe, T.M.S. (2011). Theories of learning Therese. In A. B. Nsamenang, & T. M.S. Tchombe (Eds.), *Handbook of African educational theories and practices: A generative teacher education curriculum* (pp. 175-193). Human Development Resource Centre (HDRC).

Ujah, E.U. (2020). *Development and validation of an introductory technology achievement test*. Unpublished M.Ed. Thesis, University of Nigeria, Nsukka.

Upadhya, B. & Singhlk, Y. K. (2008). *Advanced educational psychology*. APH Publishing Corporation.

Wall, D. (2010). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System*, 28(5), 499-509.

Wall, D., & Alderson, J. C. (2013). Examining washback: The Sri Lankan impact study. *Language Testing*, 10(1), 41-69.

Wikipedia.(2012). en.wikipedia.org. http://en.wikipedia.org/wiki/Jigawa_State.