International Journal of Education and Evaluation (IJEE) E-ISSN 2489-0073 P-ISSN 2695-1940
Vol 10. No. 2 2024 www.iiardjournals.org

Construction and Validation of Agricultural Science Achievement Test for Senior Secondary Three (SS3) Students in Public Schools in Delta State

Edhebe, Ikiehoke John, Prof. P. U. Osadebe and Dr. P. A. U. Ossai
Department of Guidance and Counselling, Delta State University, Abraka

DOI: 10.56201/ijee.v10.no2.2024.pg152.160

Abstract
The study examined the construction and validation of Agricultural Science Achievement Test (ASAT) for Senior Secondary Three (SS3) students in public secondary schools in Delta State. This study utilised an instrumentation research design. A total of 96 test items were created and validated using a sample of 399 students selected at random. The study was guided by four research questions. The data gathering instrument utilised was the ASAT. The participants' responses were utilised to address the study inquiries. Calculations were performed for item analysis, reliability coefficient and standard deviation. The ASAT was conducted on three separate occasions. The results of the administrations revealed that the minimum item difficulty indices rose from 0.36 to 0.40, while the maximum item difficulty indices decreased from 0.79 to 0.70. The minimum item discrimination indices increased from 0.22 to 0.32, whereas the maximum discrimination indices decreased from 0.66 to 0.44. The effectiveness of distracters increased from 0.12 to 0.17 in the first and second test administrations, but dropped to 0.11 in the third administration. On the other hand, the maximum distracter indices increased from 0.45 to 0.68. The reliability coefficients for the three test administrations were 0.92, 0.90 and 0.95, respectively, all of which fell within acceptable bounds. The study's findings indicated that the parameters of difficulty and discrimination indices, as well as the efficiency of distracters, were all within the acceptable range for a standardised test. Additionally, the reliability coefficient of the test items was high across all three test administrations. Therefore, it was recommended that educators and other individuals involved in education in Delta State and abroad should consider implementing this test.

Keywords: *Construction, Validation, Reliability, Agricultural Science, Achievement.*

Introduction

Agriculture science can be defined as the scientific study of various principles and practices related to the production, processing and marketing of agricultural products. It encompasses a wide range of topics including soil science, crop production, animal science, agricultural economics, and sustainable farming practices. The primary goals of teaching agriculture science at the secondary school level in Nigeria are to impart students with a fundamental comprehension of the concepts

and methodologies encompassing agricultural production; foster an appreciation for the importance of agriculture in food security and economic development; equip students with the knowledge and skills necessary to pursue further studies or careers in agriculture; promote the adoption of sustainable farming practices that minimize the negative impact on the environment; develop practical skills, such as crop cultivation and animal husbandry, which can be applied to real-life situations; encourage entrepreneurship and self-employment opportunities in the agricultural sector.

One way to measure the attainment of the Agricultural Science objectives is by evaluating student achievement. This can be done through assessments, examinations, class work, projects, and presentations. The performance of students can indicate whether they have acquired the necessary knowledge and skills in Agricultural Science. Tests are the predominant method of evaluating education in schools, despite the existence of other assessment instruments. Evaluations are necessary to ascertain whether students in Delta states have acquired the intended proficiency as a consequence of studying the content from the SS3 Agricultural Science curriculum. Smith and Johnson (2020) define a test as a set of questions that individuals must answer. Through these answers, examiners can determine if the test takers possess the desirable qualities that the test aims to measure. The level of Agricultural Science skill possessed by a student cannot be visibly perceived. The measurement could only be quantified through the utilisation of an Agricultural Science examination. Zhang and Kunnath (2019) argue that both standardised achievement tests and teacher-made achievement tests are designed to assess the efficacy of an adopted curriculum. According to Panadero et al. (2017), teacher-made tests may sometimes lack clarity in their objectives, or fail to effectively communicate those objectives to students. Additionally, the test items may be either excessively challenging or overly simple, and the tests may also lack validity and reliability. Nevertheless, examinations continue to be employed in Nigeria as a means of evaluating placement, ongoing assessment, prediction, and educational counselling.

Test construction and validation are crucial processes in educational and psychological research (Hambleton & Zeniskey, 2019). They involve producing a reliable and valid assessment tool to measure a particular construct or gather information about individuals' knowledge, abilities, attitudes, or behaviours. The teacher is responsible for designing the instructional plan, which includes the concepts, values and abilities that students need to learn. Additionally, the teacher designs achievement tests to assess the students' level of proficiency. A test created by the teacher is referred to as a teacher-made test. It is based on specific content areas as taught by the teacher, and is tailor-made to the teacher’s purpose.

The primary purpose of a teacher-made test is to assess the students' level of mastery in the units of instruction, evaluate the extent to which specific local objectives emphasised by the teacher have been met and serve as a foundation for assigning subject marks (Asuru 2015). However, there are some flaws in the teacher-made tests. A major flaw is incomparability of standards: The scores from one teacher-made test cannot be used for comparison with the scores of another group tested with a different teacher-made test. Mahajan (2015) suggested using standardised achievement tests to address the limitations of teacher-made tests. These tests are meticulously designed to assess objectives that are shared by multiple school systems, aiming to mitigate these downsides.

Assessments gauge understanding of information, ideas and fundamental principles. They are mostly employed to inform decisions at the classroom, state, national and international levels. These assessments are specifically tailored to align with the objectives and learning outcomes of a given subject. In a nutshell, standardized tests are valid, reliable, usable and fair. They make for objectivity, comparability and accountability (Obilor, 2017).

In order to enhance the quality and effectiveness of teacher-made tests, Mahajan (2015) proposed a set of steps to ensure their validity, reliability, usability, objectivity, comparability and accountability. These steps include: test planning, test preparation, test administration/pilot testing, final test tryout, preparation of the final test form, and establishment of reliability and validity. Test planning is a crucial stage in constructing an achievement test, particularly for standardised tests. To ensure proper planning, the researcher should consider the following factors: the individuals responsible for writing the test, the intended use of the test scores, the timing and location of the test administration and the methodology for test construction.

The process of developing a standardised test involves the creation of the test itself, as well as the formulation of the blueprint, objectives and content of the test (Pooja & Baliya, 2016). The preparation of the preliminary draft of an achievement test involves two distinct stages: item-writing and item editing. Item writing necessitates a high level of proficiency in formulating the objectives, anticipated behaviours, and constructing the blueprint that functions as a framework for creating items in the initial draft. The specific item should be documented in order to assess instructional objectives accurately. Next is the process of editing an item. Item editing involves the evaluation of items by language and subject specialists to identify any errors in language usage or unintended flaws in wording. It also aims to ensure that the items accurately assess the intended degree of achievement. The preliminary draft is revised and implemented on a selected group of students after incorporating the recommendations of the experts (Ndirika, 2012).

Item analysis is the subsequent phase in the process of test construction. Item analysis is a statistical approach used to choose suitable items for the final version while discarding inadequate ones. It involves analysing the students' replies within the sample group for each individual test item. Item analysis refers to the process of evaluating the relative difficulty index, discriminating power, effectiveness of distractors, validity check and determination of reliability coefficient of a test (Mahajan, 2015; Jayanthi, 2014).

Item analysis is the procedure of evaluating the performance of an item to determine if it accurately measures the same construct as the entire test (Obilor, 2019). Item analysis commences subsequent to the administration and scoring of the test. Obilor (2019) defines item analysis as a thorough and methodical evaluation of the testees' replies to each item in order to ascertain the level of difficulty and the item's ability to differentiate between high and low performers. The process of item analysis entails organising the scores in a descending order, starting from the greatest score and ending with the lowest score. The scores of the top 27% and bottom 27% of testees are chosen and utilised. For example, if the test was piloted with 300 students, the scripts will be organised in a decreasing order based on their scores. A total of 81 scores, representing 27% of the 300 scores,

will be chosen as the highest scores. Similarly, another 81 scores, again representing 27% of the 300 scores, will be selected as the lowest scores. The remaining scripts will be discarded.

Reliability pertains to the extent of consistency in test scores. Several techniques can be employed to estimate reliability, including test-retest, split half, alternate forms, Kudar-Richardson and Cronbach's Alpha. Validity pertains to the capacity of a test to accurately assess the specific construct it was intended to evaluate. The methods used to assess validity include face validity, content validity, construct validity, predictive validity and concurrent validity. Item difficulty, as defined by Obilor (2017), refers to the proportion of students who answer a particular item correctly. In this instance, it is also equivalent to the average value of the object. The item difficulty index spans from 0 to 100, with higher values indicating more ease of the item (question). Item discrimination pertains to an item's capacity to distinguish amongst students based on their level of knowledge in the subject matter being assessed. Ideally, the upper group of testees should have a higher percentage of right responses compared to the lower group. The success of distracters lies in their capacity to attract pupils who have misconceptions or errors in their thinking and reasoning, typically those with lower overall aptitude. It is anticipated that all the distractors in each item should function efficiently. Consequently, it is necessary for each individual in both the higher and lower groups to select at least one distracter.

Item selection is the preliminary stage of test construction that involves choosing the appropriate number of items required for the test. The items should be initially sorted in decreasing order according to their discriminative power magnitude. Individuals with values of zero and negative should be categorically rejected. Regarding item difficulty, there is no predetermined value to be chosen; rather, it is contingent upon the intended purpose of the test. A wide variety of talents or achievements is accommodated in most achievement tests by selecting difficulty levels between 40% and 60% (0.4 and 0.6). The necessity to write more items than necessary is based on the act of rejecting certain items. Content coverage focuses on the pertinence of the content of each item and as a collective entity. It also pertains to the methodical analysis of all items in the test to ascertain the extent to which the test material encompasses a representative selection of the subject matter (Asuru, 2015). The determination of the content coverage is based on expert judgment. Experts in the subject area should examine the test to see whether or not it covers all the vital contents of the subject and behaviours that need to be included in the test and whether there is a balance between topics and behaviours. Tests are usually constructed using the scheme of work for the subject and specific to the class. It is against this background this study investigated the construction and validation of Agricultural Science achievement test in Delta State.

Statement of the Problem

It has been observed that most teachers, if not all, either pick up published past questions or simply write down their own test items without considering the psychometric properties or the quality resulting from such teacher- made tests. It is known that teacher-made tests have some flaws which affect the performance of the students such as: teacher-made tests show gross neglect in item quality which truncate educational values that the subject is expected to promote in student's personal character, teacher-made tests also neglect certain prescribed areas in curriculum content

and most often, language, entry behaviour, age, sex, are not given due consideration during item construction. In a bid to overcome these problems and to improve the performance of students in agricultural that this research sought to construct and validate an achievement test on Agricultural Science for Senior Secondary (SS 3) students.

Purpose of the Study

The purpose of this study is to develop and validate ASAT for public SS3 students in Delta State. The specific objectives of the study are to:

1. construct test items in Agricultural Science for students in Public Senior Secondary Three (SS3) in Delta State.;
2. determine the validity of the test items;
3. carry out item analysis of the test (compute difficulty indices, discrimination indices and effectiveness of distracters);
4. estimate the reliability of the test.

Research Question

The following research questions guided the study:

1. What is the developed ASAT?
2. What procedures were adopted in validating ASAT?
3. What is the reliability coefficient of the ASAT?
4. What are the difficulty, distracter and discrimination indices of the ASAT?

Methods

The study focused on the construction and validation of an ASAT for Senior Secondary Agricultural Science Students in Delta State Public Secondary Schools. The population comprised 215,000 Senior Secondary Three (SS3) students. The sample size for the study, consisting of 399 JS3 students, was determined using the Taro Yamane's formula. The sampling approach used was the multistage sampling technique. Initially, the researcher employed a random sampling technique to select five public secondary schools from each of the senatorial districts, resulting in a total of 15 public secondary schools that were included in the study. Subsequently, a proportionate sampling procedure was employed to choose a sample of 399 students.

The data collection tool employed was the ASAT, a researcher-designed assessment including 96 objective test items. Each issue had five possibilities (A-E), with only one option being the right answer. The instrument's reliability coefficient, calculated using Kuder-Richardson 20 (K-R 20), is 0.86. The researcher and two research assistants administered the instrument three times consecutively, with a two-week break between each administration. Responses to the test items were scored and difficulty, discrimination and distracter indices were computed for each item. To carry out the item analysis of computation of difficulty, discrimination and distracter indices the test papers were arranged in order of size from the highest score to the lowest score, the upper 27% of the papers selected (in this case 108 papers) and the lowest 27% of the lowest scores also selected (again 108 papers), and the rest 183 papers are discarded. Finally, the highest and lowest

scores were used to compute the difficulty, discrimination, and the distracter indices (Reena & Anisha, 2017; Obilor, 2017).

Results

Research Question 1: What is the Constructed ASAT? The ASAT is a collection of multiple-choice achievement test items. The test consisted of ninety-six (96) multiple choice items. Each item has five options (A – E) with only one correct option per item. Agricultural Science Senior Secondary syllabus 2023/2024 was the main source of item in the test. The syllabus consisted of twelve (12) major content areas covering from first to third term. Each content was divided into sub-units with performance objectives expected to be achieved by students.

Research Question 2: What procedures were adopted in validating ASAT? The procedure adopted in validating ASAT was in segments. (i) Face validity and (ii) content validity. The research findings showed that the test items were 96 multiple choice objectives questions having five options (A – E) with one correct option per item. The selected students were given the test items three times in a row, with a two-week gap between each administration. Agricultural Science Senior Secondary syllabus was the main source of item in the test. Face and Content validation were carried out on the instrument. The validation of the test was done by experts in Educational Measurement and Evaluation of the Delta State University. The experts assessed the appropriateness of the items in relation to the given content area. Any item that was inadequately matched to the test blue print was dropped or revised and 60 out of 96 test items were accepted for use. Face validity ensured that the test experts looked over the test items to ascertain whether or not the test items appeared or looked like test items in Agricultural Science. The test specialists needed to examine the list of content areas in Agricultural Science, the test blueprint, and the test items that were supposed to be based on each content area in order to ensure content validity. The experts assessed whether the goods were suitable for the specified content area. Any item that did not conform to the content area and test blueprint was eliminated or modified. Editing the items in this way ensured the face and content validity with regards to the contents and the objectives measured.

Research Question 3: What are the discrimination, difficulty and distracter indices of ASAT?

Table 1

Computed Difficulty, Discrimination, and Distracter Indices for the ASAT.

Test administration	Discrimination Indices (D)	Difficulty Indices (P)	Distracter Indices (Do)
1st	0.22-0.66	0.36-0.79	0.12-0.45
2nd	0.22-0.52	0.38-0.72	0.17-0.68
3rd	0.32-0.44	0.40-0.70	0.11-0.57

Table 1 shows the breakdown of the discrimination, difficulty and the distracter indices of ASAT for first, second and third test administration. The discrimination index (D) for first test

administration was at the range of 0.22 to 0.66; for the second test administration, discrimination indices were between 0.22 and 0.52; while for the third test administration, discrimination indices were between 0.32 and 0.64. The item difficulty (P) indices for the first, second, and third test administration were 0.36 to 0.79, 0.38 to 0.72 and 0.40 to 0.70 respectively. However, distracter indices (Do) ranged from 0.12 to 0.45, 0.17 to 0.68 and 0.11 to 0.57 respectively for first, second, and third test administrations as shown in Table 1. The results showed that ASAT had difficulty indices, dissemination indices and effectiveness of distracters within the acceptable limits.

The results also showed that minimum item difficulty increased from 0.36 to 0.38 and 0.40 in the first, second and third test administration respectively and while maximum item difficulty respectively decreased from 0.79, 0.72 and 0.70 in the three-test administration. An achievement test must include varying difficulty levels and should provide reasonable opportunity of attempt to student of different intellectual levels (Sidhu, 2007). He also opined that for a five – option multiple choice objective items, the average proportion on correct responses should be about 0.69. Asuru (2015) was of the opinion that in norm-referenced test, an acceptable item difficulty range is between 0.40 and 0.60 or optimum of 0.50. Sidhu (2007) stressed that maximum reliability coefficient will be obtained if the difficulties are between chance of success and 100 percent and then recommended a difficulty level of 0.70 for 5 – option multiple choice objective test items.

According to Sidhu (2007), the discrimination index of an item indicates whether or not the item is measuring the same ability as the test measures and shows how well an item discriminates between the able and weak students. Item discrimination in this study has minimum indices increasing from 0.22 to 0.32 and maximum indices ranged from 0.64 to 0.66 in the three test administrations. Jayanthi (2014) and Sidhu (2007) were of the view that items would be considered acceptable if the facility indices were satisfactory and discrimination index 0.25 and better in excess of 0.40. The dissemination indices in this study ranged from 0.32 to 0.64 in the final test administration and is considered to be within the acceptable limit (Table 1).

Effective distracter is a plausible wrong option of multiple-choice test meant to entrap testees who do not possess the knowledge and competence being assessed (Asuru, 2015). The distracter indices for this study ranged from 0.11 to 0.68 in all three test administrations. However, all the distracters were chosen by the upper and lower group but more of the lower group than the upper group. This means that this Agricultural Science Achievement Test (ASAT) is suitable to be used by teachers to access students' performance in Agricultural Science in junior secondary schools in Akwa Ibom State.

Research Question 4: What is the reliability coefficient of ASAT?

Table 2

Estimation of Reliability Coefficient the ASAT Using K-R 20.

Test administration	No of students	No of items	Pq	$\bar{x}$	SD	SD^2	r
1st	399	96	23.06	55.54	16.83	283.60	0.92
2nd	399	96	22.16	56.00	15.00	225.60	0.90

3rd	399	96	18.24	55.10	18.39	338.00	0.95

The reliability coefficient (r) of ASAT was calculated with Kuder-Richardson-20 (K-R20) for the three test administrations and the results obtained were 0.92, 0.90 and 0.95 for first, second and third administration respectively as indicated in Table 2, yielding an average reliability coefficient of 0.92. The computed reliability coefficients for this study are within the acceptable range. The findings align with the results obtained by Reynolds et al. (2019), who developed and verified a proficiency assessment for Mathematics in secondary schools.

Conclusion

This study has constructed and validated 60 items of ASAT for use in public Senior secondary schools in Delta State. The test was duly validated with reliability coefficients of 0.92 obtained. The test possesses appropriate difficulty, discrimination and distracter indices, making it very adequate for use in Senior Secondary schools, not only in Delta State, but in all states in Nigeria and beyond.

Recommendations

To construct and validate an achievement test based on this study, the following recommendations were made:

1. Teachers should make use of this test as a guide in constructing a good quality test in other subjects.
2. The test should encompass a comprehensive selection of the topics and sub-topics addressed in the contents.
3. The writing of test items should be in appropriate language and format.
4. Procedure and principles for analysis of the test scores should strictly be followed.
5. The pool of items should undergo evaluation by another expert to identify any ambiguities and grammatical errors.
6. This test ASAT should be used in Delta State and all other schools where Agricultural Science is offered with appropriate updating from time to time.

References

Asuru, V.A. (2015). *Measurement and Evaluation in Education and Psychology (2nd Ed).* Port Harcourt: Pearl, Publishers International Ltd.

Hambleton, R. K., & Zenisky, A. L. (2019). Translating and adapting tests for cross-cultural assessments and international surveys. *Applied Psychological Measurement,* 43(1), 1-18. doi:10.1177/0146621618781038

Jayanthi, J. (2014). Development and validation of an achievement test on Mathematics. *International Journal of Mathematics and Statistics Invention (IJMSI),* 2(4), 40-60.

Mahajan, G. (2015). Construction and validation of achievement test in economics. *International Journal of Humanities and Social Science (IJHSSS),* 1(6), 54-60.

Ndirika, M. C. (2012). Achievement variation of Basic Science students taught with teacher-centred, teacher/student-centred and student-centred instructions in Kaduna State, Nigeria. *Journal of Educational and Social Research,* 2(8), 1-10.

Obilor, I. E. (2019). *Essentials of measurement and evaluation*. Port Harcourt: Sabcos Printers & Publishers.

Panadero, E., Jonsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review,* 22, 74-98.

Pooja, B. & Baliya, J. N. (2016). Construction and validation of achievement test in Science International *Journal of Science and Research (IJSR),* 5(6), 1 – 4.

Reena, R. A. (2017). Construction and standardization of Mathematics Achievement Test for IXth Grade students. *International Journal of Education and Applied Social Science,* 8(2), 651-655.

Reynolds, M. R., Livingston, J. A., & Willson, V. L. (2019). Evaluating a measure of teacher knowledge: A cognitive assessment for educators (CAE) science formative pilot study. *Assessment for Effective Intervention,* 44(1), 46-57. doi:10.1177/1534508418786067.

Sidhu, K.S. (2007). *New approaches to measurement and evaluation*. New Delhi, India: Sterling Publication Pvt. Ltd.

Smith, J. K., & Johnson, L. M. (2020). The process of test development: Initial stages. *Journal of Educational Measurement,* 57(2), 189-207. doi:10.1111/jedm.12222.

Zhang, D., & Kunnath, S. K. (2019). The effects of different item types in computer-based assessment: A meta-analysis. Computers & Education, 141, 103603. doi:10.1016/j.compedu.2019.103603.