Journal of Education and Practice www.iiste.org
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol 2, No 3

Demonstration of The Multiple Matrices Sampling Technique In Establishing The Psychometric Characteristics Of Large Samples

ANIGBO, Leonard Chinaedum
Department of Science and Computer Education,
Enugu State University of Science and Technology, ENUGU, NIGERIA
E-mail: drlcanigbo@gmail.com

Abstract

The multiple matrixes sampling technique is an approach to achieving broad curriculum coverage while minimizing testing time per student. This involved developing a complete set of items judged to cover the curriculum, then dividing them into subsets and administering the subsets to samples of pupils. This method, by limiting the number of items administered to each pupil, limits the amount of testing time required, while still providing, across students, coverage of a broad range of content. In a recent study, a test battery with 127 items in three sets was developed and standardized with the aid of the multiple matrices sampling method. By adopting this method, the study has demonstrated the possibility of validating a large number of items without unnecessarily disrupting the academic program of the pupils.

1.0 Introduction

The multiple matrices sampling of items, according to Child and Jaciw (2002), is an attempt at achieving broad coverage of the curriculum during testing while minimizing the testing time per child. Sometimes, a test developer may want to develop a pool of items that may cover, say the curriculum of the whole class or a term. In such cases, the number of items will be so many that it will not be practicable to administer all the items to the same candidates at a single seating. Also there may be complaint of disruption of academic program if the developer seeks to use the same subjects for repeated administrations of separate subsets of the test. These situations call for a more pragmatic approach that will be both time and cost saving and still achieve the desired purpose. This is where multiple matrix sampling becomes indicated.

In multiple matrices sampling, a complete set of items judged to cover the curriculum are developed and then divided into subsets. These subsets are administered to different sets of testees. By this method, the number of items administered to each set of testee is reduced while still providing, across students, coverage of a broad range of content.

According to Shoemaker (1975), although each examinee tested is administered only a portion of the items, the result from each subset may be used to estimate the statistics of the universe scores which would have been obtained by administering all the items to all the examiners. Sachar and Suppes (1980) further submitted that the multiple matrix models allow for the making of the generalization about the domains of items without having to consider the whole universe of items domain. The most important gain in this procedure is that individuals are tested on only a portion of the test items in the total pool, and yet the parameters (mean of test score, variance of test score), of the universe scores can be accurately estimated.

There are some considerations to be made in the applications of the multiple matrixes sampling. These includes: the number of subsets, the number of items per subset and the number of examinees administered each subset. These variables can be manipulated to create several multiple matrix sampling plans. The design to be adopted by the test developer will depend on the situations on the ground. These may be in the form of the number of available examinees, times available and the cost of materials. The overriding consideration, though, would be the sampling plan that would yield more precise estimates of the item parameters.

Popham (1993) identified two types of matrix sampling. These are variously characterized as item sampling type and genuine matrix sampling type. The item sampling type refers to the situations where subsets of the test package are administered to every student or group of students. This can be done in one of the two ways: - (i) The students or groups are randomly assigned to take only one of the subsets of the test package. This method saves time and ensures the coverage of the entire curriculum. The only problem is that comparing results across students or groups involves some rigorous computations.

Table 1. Matrix Sampling: Item Sampling Type.

Student Or groups		Set 1	Set 2	Set 3	Set 4
		Xx			
			Xx		
				Xx	
					xx

(ii) The second method is what Childs, Dings, and Kingston, (2002) called, the Partial Matrix Sampling. This is the situation where a subset of the package is selected to be common to all the students or groups. The remaining subsets are then matrix sampled. The common subset serves as an anchor and helps to improve comparability of students or group results, while the matrix- sampled items increase content coverage per testing time.

Table 2. A Partial matrix sampling design

Student or groups	Common	Set 1	Set2	Set 3	Set 4
1	xx	xx			
2	xx		xx		
3	xx			xx	
4	xx				xx

Table 3. Genuine matrix sampling design

Groups/sets	Set 1	Set 2	Set 3	Set 4	Set 5	Set 6
1	xx		Xx			xx
2	xx	Xx			xx	
3		Xx		Xx	xx	
4	xx			Xx		xx

5		Xx	Xx		xx	
6			Xx	Xx		xx
7	xx		Xx		xx	
8		Xx		Xx		xx

The **genuine matrix sampling** design as illustrated in the table above refers to the situations where the students or groups of students are sampled so that only some of the students or groups take any test at all. In this case, both items and students or groups are sampled

In adopting the multiple matrixes sampling design, two gains are usually borne in mind: content coverage and testing time. There are, however, some implications associated with the design. These relate to reliability, comparability and validity of the scores generated from the subsets of the total package. Reliability refers to how accurate and how consistent scores are. Test designs yield varying degrees of consistency. Generally, a test with more items tends to yield increased reliability while shorter tests yield lower reliability. Consequently if the number of items an individual student answers is small, the reliability of the student scores will be small. However, if multiple forms of the test are administered, the number of items contributing to the students score may be large. According to Shoemaker (1971) a classical test theory analysis of the resulting data would yield a mean test score for each group of students who happened to take the same items, and the mean total score would be computed as a weighted composite of the subgroup scores. The standard error of the total test score based on the matrix test would be smaller than the standard error from a test of the same length, but in which all students scores were based on the same items. In Item Response Theory (IRT) analysis of data, however, administering different items to different students would not necessarily affect reliability of the tests. In IRT, reliability is only affected by how “adaptive” the test is. According to Childs and Jaciw (2002), a test is said to be adaptive to the extent to which it does not contains too easy and too difficult items. Too easy items would make all the testees to score it correctly while too difficult items makes all score it incorrectly. The degree of adaptiveness affects the reliability of tests under IRT and not necessarily the length of the test. Matrix sample design therefore increases reliability of the total test as it greatly expands the number of items administered to the populations. The part score reliability is however decreased when matrix samples are used.

On the other hand comparability of scores across test takers is enhanced when all students respond to the same items-unlike when they respond only to subsets. This is more pronounced in a Classical Test Theory analysis. With Item Response Theory (IRT) the items can be calibrated or equated onto a single scale and this simplifies comparability.

Validity refers to the extent to which a test is measuring what it is interested to measure. The matrix design affords the test developer the opportunity to cover more of the syllabus thereby enhancing the validity of the instrument.

In line with the above specifications, a Mathematics Achievement Test Batteries (MATBat: Anigbo, 2006) was developed and standardized using a combination of the Genuine Matrix Sampling design and the Partial Matrix Sampling design. While both the items and the testees are sampled, a subset of the package was administered to all the candidates as an anchor to enhance comparability during trial testing. For the final administration, the Genuine Matrix Sampling design was adopted. Samples of pupils were required to attempt only a subset of the final package.

2.0 Methodology

2.1 *Research design.*

This was an instrumentation study. According to Garba (1993), instrumentation studies are appropriate when introducing practices, techniques or instruments for educational practice. In this study, an attempt is being made at introducing a new set of standardized test batteries based on the Primary 4 Mathematics Curriculum Modules.

2.2 *Area of study.*

The study was carried out in Enugu State. Currently, Enugu State has a total of seventeen (17) Local Government Areas. These seventeen (17) Local Government Areas are divided into four operational zones by the State Primary Education Board (SPEB).

2.3 Population of the study

The population of the study was made up of all the 48,864 primary 4 pupils in Enugu State. This number is made up of 40,236 primary 4 pupils in all the 1,134 Government owned Primary Schools and 8,628 primary 4 pupils in all the 126 registered privately owned Primary Schools in the State. (Source: SPEB, Enugu, June 2004).

2.4 Sample and sampling procedures

A total of 900 pupils were used for the final administration of the instrument. Sampling for this number was done through a multi-stage sampling technique. First, this number was shared proportionately among the four operational zones in the State using the relative numbers of Primary 4 pupils in both public and private schools in the zones. Next was the sampling of the Local Government Areas to represent each zone. This was done through simple random sampling. Names of the Local Government Areas were written in small pieces of paper and one was picked to represent the zone. The final stage was the selection of the primary schools to be used in each Local Government Area. This also was done through random sampling for both the public and private schools.

Below is the list of the sampled Local Government Council Areas and the numbers of primary 4 pupils sampled from each zone:-

Table 4: The sampled LGA'S and the numbers of Primary 4 pupils sampled for the study.

Zones/LGA's/ School types.	**1 (Aninri)**	**2 (Isiuzo)**	**3 (EnuguSouth)**	**4 (Nsukka)**	**Totals**
Public	**129**	**198**	**153**	**261**	**741**
Private	**18**	**72**	**30**	**39**	**159**
Totals	**147**	**270**	**183**	**300**	**900**

2.5 Instrument Development.

The instrument for this study is the Mathematics Achievement Test Batteries (MTBat) developed by the researcher. It is a 135-item package drawn from the contents of the primary four (4) Mathematics curriculums as structured into modules. Five multiple-choice items were developed for each of the 27 modules from the primary four (4) Mathematics curriculums. The following steps were steps followed in the development of the instrument:

2.51 Identification of the modules in the primary 4 Mathematics curriculum that should be assessed by written work.

The guideline for the use of the modules contains some recommended assessment techniques for each module. Some modules are to be assessed through written work while others were to be assessed through practical work. Of the thirty one (31) modules in the primary 4 Mathematics curriculum, twenty seven (27) of them are to be assessed through written work. The other four (4) modules were to be assessed through practical work.

.2.52 Construction of the Table of Specifications (or Test Blueprint).

Since each of the modules is to be of one-week duration, equal weight was attached to them as to the number of items to test each module. Only three cognitive levels, knowledge comprehension and higher, as advised by other researchers (Inyang and Jegede, 1991, and Mbajiorgu, 2002), were considered. In this arrangement, the four higher-level objectives (application, analysis, synthesis and evaluation) were grouped together.

2.53 Construction of Item to Measure the Stated Objectives.

Ten (10) items were constructed for each of the twenty-seven (27) modules giving a total of 270 items. This is in line with the recommendation of Oladunu, (1996) that about twice the needed number of items be used for the tryout.

Journal of Education and Practice www.iiste.org
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol 2, No 3

2.54 Validation of the Instruments**.**

The nine subsets of the batteries along with copies of the primary 4 Mathematics curriculum modules were submitted to three experts in Science Education and nine senior primary school teachers. Each expert received the total package with the primary 4 Mathematics curriculum modules while the primary school teachers received only a subset of the package. The comments of the validations were incorporated in the review of the items for trial test and for the final package.

2.55 Trials testing.

Due to the large number of items that were involved, the **partial multiple matrixes sampling** design was adopted in the administration of the trial test package. The package was divided into nine sets of thirty items covering three modules each. In addition, one module (modules 16) was made common to all the sets yielding a total of forty items per set.

Table 5. PARTIAL MATRIX SAMPLING OF THE MODULES FOR TRIAL TESTING

Sets	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**
1	**X**														**X**	**X**			**X**								
2		**X**								**X**						**X**										**X**	
3			**X**									**X**				**X**	**X**								**X**		
4				**X**									**X**			**X**						**X**					
5					**X**											**X**											
6						**X**										**X**		**X**					**X**				
7							**X**				**X**					**X**					**X**						
8								**X**						**X**		**X**				**X**							
9									**X**							**X**								**X**			**X**

Key: x: Modules used in the set

Thirty packs were produced for each of the nine sets yielding a total of two hundred and seventy (270) packs. These were then distributed to primary school pupils drawn from three Local Government Areas in Enugu state (Udi, Ezeagu and Enugu North). These were not among the Local Government Areas sampled for the final administration of the instrument. Sets of the trial test packages were randomly assigned to each of the schools. The question papers were given to the class teachers who helped in administering them to the pupils. They were thereaftercollected for item analysis.

2.6 Item Analysis.

The items of the trial test were subjected to item analysis to ascertain the psychometric characteristic of the items: - item difficulty (item facility), discrimination indices and the distracter pattern of the options.

The uppermost 27% and the lowest 27% as recommended by Kelly (1939) were respectively grouped as high and how achievers.

Items with acceptable difficulty and discrimination indices were selected to form the final package. Particularly, the best five items of each module were chosen in line with the tables of specifications.

For the analyses of the items of Module 16, which was common to all the sets, the average score of all the groups in the module was taken. The performance of the group whose average score was closest to the overall average score was used to analyze the items of the module.

2.7 Editing of the Items.

To ensure content validity of the package and in line with the Test Blueprint, five (5) items were selected from each of the modules. This is to reflect the relative emphases placed on the modules in the curriculum (the modules were to be of one week duration). The items were divided into three parts (sets) corresponding to the 3 terms of the school year. Each set has forty-five (45) items.

2.8 Reliability of the instruments

After the trial testing and item analysis, the measures of the internal consistency of the subsets were sought. The Kuder-Richardson estimate (k-R-20) was used.

Table 10: Reliability Estimates of the subsets of the MATBat

SETS	1	2	3	4	5			
K-R 20 ESTIMATES	**.93**	**.61**	**.93**	**.76**	**.68**	**.82**	**.69**	**.83**	**.80**

3.0 Discussion

With the multiple matrix sampling technique, the psycholometric characteristics of one hundred and thirty five [135] item were established in just one test administration. Ordinarily, the validation or standardization of such a large number of items would take repeated administrations. This, according to Shoemaker [1971], is one of the gains from application of the multiple matrixes sampling technique. During the trial testing of the items, a total of 270 items were administered in nine different matrixed sets. Each set was administered to 90 pupils. In the final test administration, the package was reduced to 3 sets. Where it not for the adoption of the multiple matrixes sampling technique, it would have taken nine different administrations for the trial test items and three for the final administration. This, no doubt, would have meant a major disruption of the academic programme of the schools. Bunda [1973] agrees that the adoption of the multiple matrix sample technique would result in the reduction of testing time and the cost of test administration in schools. Any intelligent school administrator would be reluctant to allow the academic program of the school to suffer repeated disruptions as a result of test development. Moreover, schools are more likely to participate when intrusions into regular activities are minimized. For this reason, test developers had resorted to reducing the number of items in their test packages. This has serious implication on the content validity of such test packages. Using the multiple matrix design for this study, the contents of primary 4 Mathematics were covered in a single test administration!

References

Anigbo, L. C. (2006), Development and Standardization of Mathematics Achievement Test Batteries (MATBat) for Primary Four Pupils in Enugu State. *Unpublished Ph. D Thesis.* University of Nigeria, Nsukka.

Bunda, M. A. (1973), An Investigation of an extension of Item Sampling which yields Individual Scores. *Journal of Educational Measurement* 10(2) 117-130.

Childs, R. A, Dings, J., R., & Kingston, N (2002), *The effect of Matrix Sampling on Student Score Comparability in Constructed – Response and Multiple- Choice Assessment.* Washington DC: Council of Chief State School Officers.

Childs, R. A & Jaciw, A. P. (2002), "Matrix Sampling of Items in Large – Scale Assessments" – A paper presented at the symposium on Provincial Testing in Connection Schools. Research, Policy and Practice, Victoria, BC. Canada.

Garba N. L. (1993), Development of an Instrument for Evaluating Practical Projects in Woodworking. *Unpublished Ph. D Dissertation.* University of Nigeria, Nsukka.

Inyang, N. & Jegede, O. (1991), Development, validation and standardization of integrated science achievement tests for junior secondary schools. *Journal of Science Teachers" Association of Nigeria,* 27 (1): 21-29.

Kelly, T. L. (1939), The selection of upper and lower groups for the validation of test". *Journal of Educational Psychology*. 30:17-34.

Mbajiorgu, N. M. (2002), Effect of science-technology-society approach on scientific literacy and achievement in Biology. *Unpublished Ph. D. Dissertation*. University of Nigeria Nsukka.

Oladunu. M. O. Ajayi M. A. and Ogumbayo F. O. (1996), *Test*, *Measurement and Evaluation in Education.* Ibadan: Demolade Omotayo Publishers.

Popham, W. J. (1993), "Circumventing the High cost of Authentic Assessment". *Phi Delta Kappa*. 7,47,473

Sachar, J & Suppes (1980) "Estimating Total Test from Partial Scores in Matrix Sampling Design". *Journal of Educational and Psychological Measurement.*40 (1) 687-699

Shoemaker, D. M. (1971), *Principles and Procedure of Multiple Matrix Sampling* (Technical Rep. 34). Inglewood, CA: Southwest of Regions Laboratory for Educational Research Development

Shoemaker, D. M. (1975), "Application of Items Banking and Matrix Sampling to Educational Assessment." In Derringer. D. N and Vanderkamp. L. T. (Ed). *Advances in Psychological and Educational Measurement*. London. Wiley Press.