THE CONSTRUCTION AND VALIDATION OF MATHEMATICS ACHIEVEMENT TEST IN PRIMARY 3 IN ZARIA LOCAL GOVERNMENT AREA.

BY

YAHAYA KORAU

DEPARTMENT OF EDUCATION
AHMADU BELLO UNIVERSITY
ZARIA.

1988

DECLARATION

I hereby certify that this thesis has been written by me and that it is a record of my own research work. It has not been presented in any previous application for a higher degree. All quatations are indicated by quotation marks or indentation and the sources of information are specially acknowledged by means of references.

CERTIFICATION

The undersigned certify that they have read and have recommended to the Faculty Board of Education for acceptance of an independent study entitled "The Construction and Validation of Mathematics Achievement Test in Primary 3 in Zaria Local Government Area" submitted by YAHAYA KORAU in partial fulfilment of the requirement for the degree of Masters of Education.

Dr. A.A. Sambo
Chairman, Supervisory Committee — 8/12/88 Date

Dr. Z.K. Oguntebi
Member, Supervisory Committee — 21/12/88 Date

Dr. A.I. Ikeotuonye
Member, Supervisio y Committee — Date

Dr. F.O. Ogunlade
Head, Department of Education — 22/12/88 Date

Prof. D.I. Saror,
Dean, Postgraduate School. — 9-2-89 Date

DEDICATION

This work is dedicated to my late mother Maimunatu Abdullahi who took the trouble of bringing me up

ACKNOWLEDGEMENTS

There was no doubt that this research afforded me with an opportunity to know alot about the primary schools in Zaria Local Government. I was made to know alot of their problems which ranges from both human and materials.

I was assisted by numerous individuals and organizations in the course of this study. I wish to personally thank a few of them that readily come to mind.

I wish to acknowledge indebtedness and gratitude to Dr. Abdussalami Abubakar Sambo of the Department of Education, Ahmadu Bello University, Zaria, who despite his numerous responsibilities provided the supervision, inspection, encouragement and other assistances that brought this work to fruition.

I am also grateful to Dr. Lawrence Shirley and Dr. Z.K. Oguntebi of the Mathematics Education Section of the Department of Education for their immeasurable contribution. I am highly indebted to Mal. Idrisu Umaru Jahun, Dr. A.I. Ikeotuonye who also spared their time to offer their professional advices which really helped me to get through with my work.

This acknowledge would not be a complete one if I did not extend my thanks to Local Education Authority Officials particularly, M.M. Dalhatu who provided me with alot of

▼

informations and data about the educational resources in the Local Government. I wish to thank all the Headmasters of the Schools that took the examination for their co-operation and kind gestures throughout the period of the study.

I am grateful to the Department of Education, Ahmadu Bello University, Zaria for its material support.

I wish to extend my thanks to all the staff of Nigerian and African Languages for their effort to translate the English version of the test into Hausa. I am appreciative to Mr. Benedict Sheyin of the Department of Education, Ahmadu Bello University, Zaria, who accepted the task of typing and retyping the drafts. His untiring efforts made it possible to come up with a neatly bundle of typed documents. His assistance is highly appreciated.

Finally, I must express my gratitude to my wives, Hafsatu and Rahinatu, who endured my absence and for taking care of our children at the time I was deeply involved in the study.

ABSTRACT

Well constructed test items for the purpose of testing mathematics achievement is an essential requirement for the understanding of the subject and its subsequent application to classroom situations. However, the main objective of this study was to construct and validate tests on primary three mathematics curriculum.

In order to accomplish this objective, three sets of parallel tests were constructed and administered to 363 class three pupils for the first try-out. It was on this basis item analysis was conducted and consequently resulted in coming up with two sets of parallel tests which were finally administered to 772 class three pupils in 20 primary schools randomly selected in Zaria Local Government Area of Kaduna state. An evaluation model was developed for assessing achievement in primary three mathematics which was used to construct the test items.

The study revealed that the mean performance of urban schools is higher than the mean performance of rural schools in each test. The result of the finding also showed that the mean performances of boys and girls were relatively the same in test in all the tests.

The reliability coefficients for all the variables in the study were significant at 0.05 level of significance. At the same time the validity coefficient for test I and test II were found to be .88 and .86 respectively.

TABLE OF CONTENTS

		Page
Title Page	.....	i
Declaration	.....	ii
Certification	.....	iii
Dedication	.....	iv
Acknowledgement	.....	v
Abstract	.....	vii
Table of Contents	.....	viii
List of Tables	.....	xii
List of Appendices	.....	xiii
Definition of Terms	.....	xiv

CHAPTER 1: THE PROBLEM

1.1	Introduction to the Study	.....	1
1.2	Purpose of the Study	.....	4
1.3	Significance of the Study	.....	5
1.4	Statement of the Problem	.....	8
1.5	Research Hypotheses	.....	9
1.6	The scope and Delimitations of the Study	.....	9
1.7	Summary	.....	10

CHAPTER 2: REVIEW OF RELATED LITERARURE

2.1	Introduction	.....	12
2.2	The Concrete operational State	.....	13
2.3	Mathematics Educator's Comments on the Topics of Primary 3 Mathematics Curriculum	.....	16
2.4	Concepts of Evaluation	.....	18
2.4.1	Evaluation and Curriculum Development	.....	23
2.5	Cognitive Domain of Educational Objectives	.....	24

		Page
2.6	Evaluation Procedures	28
2.6.1	Formative Evaluation	28
2.6.2	Summative Evaluation	30
2.6.3	Norms and Criterion Referenced Measurements	31
2.7	Evaluation Instruments	33
2.8	Types of Tests	34
2.9	Planning the Test	40
2.10	Writing the Item	41
2.11	Pretesting the Items	41
2.12	Validation Process	42
2.12.1	Preparing the Final form	42
2.12.2	Difficulty Index	43
2.12.3	Descrimination Power	45
2.12.4	A look at the psychometric Analities	46
2.12.5	Validity	46
2.12.6	Reliability	49
2.12.6.1	Stability	50
2.12.6.2	Equivalence	50
2.13	Studies Outside Nigeria	51
2.14	Studies Within Nigeria	54
2.15	Sampling Techniques	56
2.15.1	Random	57
2.15.2	Stratified Sampling	57
2.15.3	Cluster Sampling	58
2.16	Summary of the Literature Survey	58

CHAPTER 3: RESEARCH METHODOLOGY

3.1	Introduction	61
3.2	The population under study	62
3.3	Sample for the First Try-out	62
3.4	Sample for the Final Try-out	65
3.5	Items Distribution	66
3.6	Item construction	72
3.7	First Try-out	73
3.8	Statistical Analysis for the First Try-out	74
3.9	The Administration of the Final Test	82

Table of Contents (Cont'd) Page

3.10	Procedures for Administering the Final Test	83
3.11	The Statistical Analysis of the Final Test	84
3.11.1	Means and Standard Deviations of the various variable in the study ..	85
3.11.2	Discrimination Inde and Difficulty levels	85
3.11.3	Statistical Design for the Data Analysis	85
3.12	Reliability Estimates	87
3.13	Standard Error of Measurement	87
3.14	Concurrent Validity	88
3.15	Summary	88

CHAPTER 4: DATA PRESENTATION AND ANALYSIS OF DATA

4.1	Introduction	89
4.2	Testing for Equivalence by the use of Means and Variances of the Two Tests Taken by the same Group ..	89
4.3	Discriminative Index and Difficulty levels	91
4.4	Comparative Analysis of the Results by Location	92
4.4.1	Comparison of the Results by Locations	92
4.4.2	Performance Between Urban Schools and Rural Schools	93
4.5	Comparative Analysis of the performance of Urban and Rural Boys; Urban and Rural Girls	96
4.5.1	Comparison Between Boys from Urban Schools and Boys from Rural Schools in Test I	97
4.5.2	Comparison Between Girls from Urban Schools and Girls from Rural Schools in Test I	97
4.5.3	Comparison Between Boys from Urban Schools and Boys from Rural Schools in Test II	98

4.5.4	Comparison Between Girls from Urban Schools and Girls from Rural Schools in Test II	99
4.6	Reliability Estimates	103
4.7	Standard Error of Measurement	104
4.8	Concurrent Validity	105

CHAPTER 5: SUMMARY, FINDINGS AND RECOMMENDATIONS

5.1	Introduction	107
5.2	Summary	107
5.2.1	Purpose of the Sgudy	107
5.2.2	Related Literature	108
5.2.3	Data Collection procedures	109
5.2.4	The Hypotheses and Instrumentation	110
5.3	Findings and Conclusions	110
5.3.1	Hypothesis I	110
5.3.2	Hypothesis II	111
5.3.3	Hypothesis III	111
5.4	Reliability Coefficients	112
5.5	Standard Errors of Measurements	112
5.6	Validation Process	114
5.7	The use to be made of VANAT	116
5.8	Recommendations for Further Research	117

BIBLIOGRAPHY	.....	120
APPENDICES	.....	126

LIST OF TABLE

Table		Page
3.1	Schools selected for the first Try-out in Zaria Matripolitan	64
3.2	Guideline for constructing the Test Items for Validation	69
3.3	Detailed Table of Specification for Test Validation for Primary 3 Mathematics Achievement Test	71
3.4	Means and Standard Deviations of the Test Sets	74
3.5	Means and Standard Deviation of the Schools selected for the First Try-out	75
3.6	Means and Standard Deviation of the Sex	76
4.1	Means and Standard Deviations of the Two Tests taken by the same group	90
4.2	T-Test of the Tests Sets	90
4.2.1	T-Test for standard Deviations	91
4.3	Means and Standard Deviations of Scores of Urban and Rural Schools	93
4.4	Means and Standard Deviations of Boys and Girls who participated in the Examination	95
4.5	Means and Standard Deviations of Sex by location	96
4.6	T-Test of Sex Differences on The VANAT; Urban Boys and Girls	102
4.7	Equivalent Reliability Coefficients of each Test by Sex and Creation	103
4.8	Standard Errors of Measurement	105
4.9	Concurrent Validity Coefficient of the Test Sets	106

LIST OF APPENDICES

			Page
APPENDIX	A:	Outline of "Primary Education Mathematics Curriculum"	127
APPENDIX	B:		127
APPENDIX	B:	First Try-out of Primary 3 Mathematics Test	128
APPENDIX	C:	First Try-out of Item Analysis	138
APPENDIX	D:	The English version of Final Try-out of Primary 3 Mathematics Test	139
APPENDIX	E:	The Hausa Version of Primary 3 Mathematics Test	149
APPENDIX	F:	Final Try-out Item Analysis ..	159

ABBREVIATION AND DEFINITION OF TERMS

Cognitive Achievement Test:	Test that is used to assess the intellectual and academic progress of the students.
DAMAT	Diagnostic Mathematics Achievement Test.
DAT	Differential Apptitude Test.
IRR	Item Retention Range.
Rural Public Schools:	Primary Schools in the rural areas that are control by the government.
SAT	Standardized Achievement Test.
Urban Public Schools:	Primary Schools in the urban areas that are control by the government.
Validation	The developmental processess of Validating a Test.
VAMAT	Validated Mathematics Achievement Test.

Chapter 1

THE RESEARCH PROBLEM

1.1 Introduction To The Study

In Nigeria today, much emphasis is placed on technological development. This has resulted in our educational system shifting emphasis from learning of Arts and humanities to science and technology. Mathematics is regarded as a service subject to Science and Technology as shown in professions such as engineering, medicine, physics and other related desciplines. Hence, mathematics is absolutely essential in the advancement of science and technology.

Mathematics teaching and learning in primary schools is often said to be going in a downward trend. This is said to be due to the simple fact that research work in the field of mathematics education is not passed on to the primary school teachers.

Many teacher training programmes assume that a teacher has acquired the method of teaching and is able to pass the required mathematics courses for certification, he is ready to impart his knowledge to his students. Hardly any emphasis is given to the techniques of evaluating mathematics achievement of the students. The reality however, is that a teacher should be able to construct and validate good

mathematics test items for his students. It is only through this he would be able to identify areas of weaknesses and strengths of his students.

Experience has shown that pupils at primary schools may acquire knowledge of Mathematics but the problem is that these pupils have not been exposed to the right kind of assessment procedures that will make them express what they know.

Perhaps this state of affairs is brought about by the problems associated with testing in Nigeria. Two major problems of testing can be identified; First there is the problem of qualified personnel. Olomolaiye (1987) reported that as of 1983 only two universities, Ibadan and Jos offer Master's degree programme in test and measurement. The second problem is that most of the tests being currently used in Nigeria measure only cognitive domain, most of which are foreign or at best adaptation of foreign tests.

In order to help solve the problem of test construction and test interpretation in secondary schools, the Department of Education, Ahmadu Bello University, Zaria, is currently sponsoring a project on the construction and validation of Achievement Test for use in Demostration Secondary School. This project if completed would come up with a well

constructed and validated test items which the students can use at any time. It is because of this initiative taken by the Department, the researcher feels it is good to embark on an independent study for use in primary 3 in Zaria Local Government Area.

Marshell and Hales (1971) observed that most teachers in America have received little or no training in the technique of test construction. Prospective teachers of elementary schools in Nigeria occasionally obtain a fleeting glance of test and measurement in principle and method class. Even then however, the course is concerned with acquenting, the students with the definitions and types of tests.

It is generally recognized by educators that provision in instruction should be made on the basis of abilities and achievement of the students. Since instructional content and achievement of students are continually changing up-to-date measures of achievement should be obtained frequently. Good measurement procedures are therefore crucial when working with individual differences because it helps to expose students' problem academically. One of the ways to find these differences is to give the individual a test, so that one could see and identify students' strengths and weaknesses.

1.2 Purpose Of The Study

The purpose of this study was to construct and validate tests battery for primary 3 mathematics content which would consequently give birth to a battery of achievement test called Validated Mathematics Achievement Test (VAMAT) for use in primary 3.

The content of the elementary school mathematics which to most individuals means arithmetic is not affected seriously by social and scientific changes as is the case with certain subjects. Mathematics is based upon fixed system of numbers that bear definate relationship to each other.

In order to solve the problem of test construction and test interpretation in primary schools, the study was therefore set to achieve the following objectives:-

1. To construct three parallel test papers based on the contemporary primary 3 mathematics curriculum.
2. To try out the tests with a reasonable number of candidates under the study for the purpose of a pilot study and developing psychometric qualities of the tests.
3. To revise and edit the test items based on the data obtained in 2 above.

4. To try it again to a larger sample as a representative of the population under this study with the view to finding acceptable psychometric characteristics of the test items.

5. To determine both validity and reliability of the tests.

1.3 Significance Of The Study.

The Federal Republic of Nigeria on the 12th March (1979) introduced new Mathematics Curricula to serve primary, teacher education, Junior and Senior Secondary Schools in the country. This is in accordance with the National Policy on Education which calls for an intensive research work at the primary school level which constitute the nucleus of the nation education system.

Also the National Policy advocates for the use of continuous assessment in our school system. One of the implications of this is that teachers are expected to use their own tests in conjunction with the published ones where available. It is in line with this, a research like this is worth doing because it will provide a pool of test items which could be used at the primary level.

Also looking at the new National Policy on Education one would see that a study like this one is necessary because of the constant involvement of teachers in determing future success of their pupils. The importance of the study is therefore dictated by the content of the National Policy on Education (1981) which introduces the 6-3-3-4 system of Education. In particular and of interest ot this study, the policy stipulates that continuous assessment procedures should be given the required attention they deserved. It is viewed as important and necessary to assess the level of mathematics required by pupils at the lower classes of primary schools which will serve as a foundation for the further work in higher classes.

Mehrens and Lehmann (1973) asserted that too many significant educational policies are being shaped as a consequence of test result, too many decisions about educational effectiveness of the teachers are being based on pupils test scores. Too many citizens are forming their views about education as a consequence of published test results. This seems to be the same in Nigeria where lots of decision are being taken based on pupils test scores. They therefore suggest that teachers must become not only knowledgable regarding test construction in general but also play an important role in shaping the nature of educational tests.

The psychological justification for choosing class 3 in the primary is borne from the fact that the behaviour patterns of children at this period proceed from concrete operational to formal operational stage. In this period of mental development the child is capable of carrying out activities that make him suitable for this kind of research. The child at this stage has passed through primary one and primary two and is assumed to have gained lots of ground in elementary mathematics.

At class 3 the child should be able to count, add, subtract, multiply and can identify some geometrical shapes. It can be seen that he is well versed in handling elementary mathematics and is now ready to go to more complicated elementary school mathematics which requires comprehension and application. It is therefore a logical decision to consider primary 3 as a transitional stage in primary education because it is the class from which the child is moving to a more difficult part of primary mathematics. The researcher feels that a research like this is worth doing because it will give a clue as to how much mathematics the children acquired before going to higher classes of primary schools.

Thus the emphasis here to diagnose the problems the pupils have in mathematics at this early stage of primary school.

Since the study will make the researcher to have a direct contact with the primary school teachers, this will definately help to create awareness in the minds of the teachers.

The study will give birth to a set of achievement test called Validated Mathematics Achievement Test (VAMAT) for use in the lower classes of primary schools. Thus tests of educational attainment will be constructed in all the mathematics topics that are taught at class three of primary schools. It is hoped that when the project is completed, the test series will be in use in other primary schools in Kaduna State of Nigeria in general.

1.4 Statement Of The Problem.

The role that construction and validation study in evaluation study plays, especially in the development of item analysis justifies the need to give a renewed and adequate attention to it in our test and measurement procedures. To that effect, this study therefore attempts to find answers to the following questions:-

1. To what extent are the two tests of Validated Mathematics Achievement Test (VAMAT) equivalent?
2. To what extent are the tests valid?
3. To what extent are the test reliable?

4. Do pupils from urban schools perform better than children from rural schools in VAMAT?

5. Do boys in primary 3 from Zaria Local Government Area perform better than girls in VAMAT?

1.5 Research Hypotheses

The following are null hypotheses to be tested in the second try-out:-

1. There is no significant difference in the variability of scores of the students who took both forms of the tests.

2. There is no significant difference in mathematics achievement between the children from Urban Schools and children from Rural Schools in each test of the VAMAT in Zaria Local Government Area.

3. There is no significant difference in mathematics achievement between boys and girls in each form of the VAMAT in Zaria Local Government Area.

1.6 The Scope And Delimitation Of The Study.

The study was carried out in Zaria Local Government Area mainly because of the proximity for the researcher. This would help the researcher to conduct the study with a minimum difficulty.

The part of the curriculum under study is restricted to the primary 3 mathematics curriculum only. The study would also be limited to only those primary schools that have mathematics teachers continuously in the Local Government Area.

As pointed out earlier, the research involved only the construction and validation of the test items for the use of primary 3 pupils within Zaria Local Government Area. The number of items in each test would be 30 because of the levels of educational and physical development of the population under the study.

The result of this would only be applied directly to pupils in Zaria Local Government. Any generalization attempted for other pupils will only be considered as a rough extrapotation of the data obtained.

1.7 Summary

This chapter mainly examined the following subheadings; The researcher brings to light the introduction to the study in which the problems facing tests and measurements in Nigeria were mentioned. The objectives of the research were outlined. The chapter also looked at statement of the problem which talked about the questions the research hopes to answer. And this consequently led to the formulation of some hypotheses for the research.

The significance, scope and delimitations of the study are well articulated and this gives a reader an insight into the justification of the research and the area which it hopes to cover.

Chapter 2

REVIEW OF RELATED LITERATURE

2.1 Introduction

Since the research concerns the evaluation of pupils at the lower level of primary school, the researcher feels it is desirable to look at what developmental psychologists and mathematics educators say at this level of education. The purpose however, is to see what they view the teaching of mathematics to young children particularly those at the concrete operational stage should be. It is hoped that this will help a reader to see how psychologists and mathematics educators handled the issue of mathematics at the early stage of the child development.

After this, the researcher concentrated his review on evaluation. Here the review centred on the following:-

1. The concept of evaluation
2. Evaluation model
3. Evaluation procedures
4. Evaluation instrument.

Having reviewed the above aspects the researcher then looked at psychometric qualities, empirical studies and sampling technique.

2.2 The Concrete Operational Stage

Although Piaget (1959) believed that there are three main stages of cognitive development namely pre-operational, concrete operational and formal operational stage, the one that is most relevant to this stage is the concrete operational stage. It is certainly the stage that coincides with the children's cognitive development the researcher is using as his population. Piaget and Inhelder (1959) asserted that the period consists of preparation and realization of concrete operations.

According to Piaget reversibility is the most defined characteristics of this period. At the early stage of this developmental period the child may be able to percieve two rows of equal numbers of coins as being the same if put into one-to-one correspondence. The child at this stage can handle same mathematical tasks the children in pre-operational stage would not be able to handle. At this stage however, the child is capable of manipulating some mathematics which Piaget says every average class three child is expected to be able to handle.

Nevertheless thinking at this level of concrete development according to Piaget shows some limitations.

> "There appear in the children's difficulties in dealing with verbal problems, the attitudes to rules and belief about the origin of objects and of names, their procedures by trail and error then by testing hypotheses in solving problems, inability to see general rules or to accept assumptions and failure to be beyond the data provided or imagine new possibilities or explanation".

This assertion is further clarified by an investigation conducted by some students, organized by lecturers through Mathematics section at the University of Miani. This experiment showed that about 40 percent or more of older Juniors in an English primary school still failed in some item testing understanding conservations of quantities.

In a nutshell what Piaget is advocating here is that normal child at this level of development needs more carefully structured activities involving more complex relationship than do infants in subjects with science biad. Below (1964) observed that curriculum designed to match children's developmental level should not expect them to engage in activities that they are not yet fully able to understand. Thus for example, one should not teach anything about addition until basic number concepts are well established and until laws of commutativity and associativity are understood. No should one teach measurement until conservation of length is established.

From what was said above, we could see that Piaget had suggested that development involves successive reconstructing of facts and relations which form children's interraction with active manipulation of their environment. Bruner (1963) on his own part focussed on just how the result of such interraction were represented in the child's mind. Thus Bruner (1963) said:

> If we are to benefit from contact with recurrent regularities in the environment, we must represent in some manner. To dismiss this problem as "mere memory" is to understand it. For most important thing about memory is not storage of past experience, but rather the retrieval of what is relevant in usable form. This proceed so that it may indeed be relevant and usable in the present when needed. The end product of such a system of coding and processing is what we may speak of as a representation.

Bruner went on to say that what the child is able to do is to give structure to things he counts, but he is not yet readily able to deal with possibilities not directly before him. The implication of this therefore, is that it is only when the child is at the stage of concrete operation he is capable of understanding a great deal of the basic ideas of mathematics. However, the child can only do such tasks that his mental development can contain.

2.3 Mathematics Educator's Comments On The Topics Of Primary 3 Mathematics Curriculum.

Number:- Commenting on early experience of children with numbers. Fakuade (1981) asserted that "Ability to repeat or say number names in order does not necessary imply understanding or knowledge of numbers". This means that when the pupils come to school for the first time the need to remember the informal ways in which children acquire number ideas at home before coming to school must be taken seriously.

Addition:- Some early research efforts focussed upon the relative difficulty of addition facts when the drill method of teaching was popular, Mactatchy (1944) and Swenson (1956) said that the difficulty always appear to be at least in part of the teaching method and suggested need to reconsider difficulty level in terms of the more meaningful methods.

Subtraction:- Gibb (1956) found that the highest degree of pupils attainment in subtraction was on take-away problems, while the lowest was on comparison problems. Nevertheless for introductory work, Obsborne (1966) reported that using set-partitioning-without-removal resulted in greater understanding than take-away problems. In another study conducted in America it was found that the decomposition method when taught meaningfully was most successful.

Multiplication:- Harvey (1966) compared the equal addition approach with the use of cartesian products. She reported that for her second grade subjects equal addition multiplication problems were less difficult to solve. At the third grade level Gray (1965) investigated achievement of pupils who used array representations exclusively for their introductory work with multiplication and found that it was very successful. Studying the use of mathematical properties to teach multiplication Hall (1967) found that stress on the commulative property was effective.

Division:- Gibb (1956) compared the subtractive algorithm of division with the distributive algorithm and suggested that advantages of the subtractive algorithm were; greater transfer to unfamiliar division situations, less difficulty for children of low intellectual ability.

Fractions:- Gunderson and Gunderson (1957) suggested that second graders could understand fraction only when they used manipulative materials. Many studies support the importance of using meaningful methods of work with fraction Capps (1963) studied the use of common denominator method for devision with fraction as compared with the inversion method. Immediate result showed these methods to be equally effective, however, retention test favoured the common denominators method.

Some earlier studies were concerned with the specific errors made by pupils in lower primary school in dealing with fraction. Major errors reported by Gossinickle (1983) for all operations were;-

a) Difficulty with renaming in simplest form.

b) Inadequate understanding of the process involved.

c) Computational errors.

Scott (1962) found that fifth-graders made more errors in fraction subtraction involving renaming and suggested that current emphasis on decimal system resulted in pupils tending to relate the process to the decimal number.

2.4 Concepts Of Evaluation

The meaning of evaluation is still relatively new to education. The term has been used to include appraisal of the school programme, curriculum and instructional materials, appraisal of the teacher of the school.

From the above it implies that evaluation is a relatively new technical term, introduced to designate a more comprehensive concept of measurement than is implied in convectional test and examination. Thus Bormuth (1970) said:

> "The emphasis in measurement is upon single aspects of subject-matter achievement or specific skills and abilities, but --- the emphasis in evaluation is upon broad personality changes and major objectives of an educational programme. This include not only subject matter achievement but also attitudes, interest, ideals, ways of thinking work habit and personal and social adoptability".

However, another school of thought views evaluation in a different angle. To them, there is a confusion concerning the meaning of evaluation as it applies to classroom instruction. In some instance it is used as a synonym for the term measurement. Thus when teachers administer achievement tests they might say earlier that they are "evaluation" achievement, with little regard for specific meaning of the two terms. In other cases evaluation is used as a collective term for those appraisal method that do not depend on measurement. Thus the use of the two terms distingueshed "evaluation as qualitative description of pupils behaviours" from "measurements" which are quantitative descriptions. Looking at evaluation from the instructional point of view Gronlund (1981) defined evaluation as "a systematic process of determining the extent to which instructional objective are achieved".

We have to note here that there are two important aspects of this definition. Firstly it shows that evaluation is a systematic process, which omit a causal uncontrolled observation of students. Secondly, it assumes that instructional objectives have been previously identified.

By and large, this definition indicates that evaluation is a much more comprehensive and inclusive term than measurement which is limited to quantitative description of students; that is the result of measurement are always expressed in number (e.g. "Shehu correctly solved 10 of the 15 arithmetic problems"). It does not include qualitative description (e.g. "Shehu's work was fine"), nor does it imply judgement connecting the worth or value of the result obtained.

The main emphasis in classroom evaluation is on the extent to which learning outcomes are achieved. It is however, as a result of this the Association of Supervision and Curriculum Development in 1967 writes that "the broader terms of the learners development as a person, is essentially an evaluation that helps him steadily towards a valid and healthy image of himself. It is therefore important for him to learn about his strengths and resources". Also commenting on the importance of evaluation to our educational system Adams and Tergerson stated (1966) said that:-

> Every teacher would soon be teaching with greater precision, because he would be getting data constantly on what was "working" and what was not, what was already mastered and what remained to be learned. Meanwhile the student himself would be channeling for he too would have a clear sense of direction.

Commenting on the same issue, Popham (1981) said when a student is transfered to another teacher or another school, the new institution or teachers could pick him up smoothly and move forward with for they would have ways of knowing the things that really matter about him. With the help of evaluation parent could be helped to become counsellors to their children because the school could help them towards that line. Employers would make use of evaluation greatly because they can select their employees from the data obtained from evaluation. From this we can see that the public and teaching profession would be in a better position to make wise decision about changes in the educational programme as a whole.

Perhaps the greatest gains of all occur within the learner himself. As he learned not only to see his academic progress with clear eyes but also to appraise himself validly. He would be more in position to take charge of his own life space and guide his own learning.

All that has been said is another way of saying that there is a great deal more to evaluation than one can think of. And that is why Furst (1976) stated that:-

> All too commonly it has been seen as if it were only measurements-a way of getting at marks and grades and credits. But when we view evaluation as feedback we instantly see how it resonates back through every stage of educational process, and this brings to light all what we wanted to know about a child's progress.

Good evaluation has a way of encouraging thought about basic purposes, values and goals. If evaluation is to be properly handled through an interactive process it would have a renewing effect upon learners, teachers, educational planners, and the public alike. Thus it will help everybody who is involved to think more about what he is after and how he is getting along. In the long run, then, a high-quality evaluation programme could be regarded as the surest way a learner, a teacher or a school system can have to obtain valid objectives.

Evaluation plays an important role in the school programme. Gronlund (1981) observed that evaluation contributes to the teaching learning process used in classroom instruction in programmed instruction, curriculum development, marking and reporting guidance and counselling, school administration and research.

2.4.1 Evaluation And Curriculum Development

On a broader scale, evaluation of curriculum itself also plays an important role in curriculum development. This is partinent because during the early stages, when new methods and materials are being tried evaluation data enable the curriculum developer to determine the effectiveness of the new methods and to identify areas where revision is neede. Ebel (1979) said when new curriculum programme has been fully developed, evaluation data makes it possible to determine the dgree to which the new curriculum is effective in meeting the instructional objectives for which it is designed.

Also citing some uses of evaluation Hammil (1971), said evaluation can be seen as reporting of pupils progress results to parents. Here one can see that the systematic use of evaluation is classroom provides the teacher with an objectives and comprehensive picture of each pupil's progress.

The result of evaluation procedures are especially useful for guidance and counselling and guiding them in the selection of curriculum and extra-curricular activities and also helps pupils to solve personal and social problems.

A comprehensive continuous evaluation programme in schools aids the school administrator. This is because from the data

collected, it is possible to judge the extent to which the objectives of the school are being achieved. It helps to identify the strength and weakness in the curriculum. Evaluation also provides the information on which to base administrative decision concerning the placement, grouping and promotion of pupils. Talking on the same issue Gerberich (1962) said that evaluation data can also play an important role in the school research programme.

2.5 Cognitive Domain Of Educational Objectives

Experience have shown that most teachers in Nigeria evaluate their students mostly in the cognitive domain. Even though this research concerns itself with only cognitive domain the researcher feels this is done at the detriment of the other domains.

Indeed, the psychologist who first thought of area of learning that need to be evaluated was Bloom. Bloom et al (1956) have classified the cognitive domain of educational objectives into six sub-domain in increasing order of complexity as follows:- Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation.

Generally speaking, cognition is referring to the students intellectual abilities in the various school subjects. Some people have tried to bring out sharp distinction between the

first class, knowledge and others by classifying the domains into knowledge and higher intellectual skills and abilities. However, Ohuche (1982) commenting on this issue said "There have also been some other variation used by specialist who have constructed questions for assessing cognitive outcomes in Mathematics".

Another important issue to agree on from the start is how to formulate our objectives. "Without getting involved in the debate which surrounds the pros and cons of the use of behavioural objectives, I will start right away that we are convinced that in cognitive functioning behavioural will satisfy our requirement"; Onwana (1980).

According to Bloom's six categories of cognitive domain of educational objectives:- knowledge refers to the level of objectives of cognitive domain which emphasize remembering of all types of either recognition or recall of data facts, trends of events, terminology, rules formulea and theorems.

The second level in Bloom's taxonomy is that of that of comprehension. Objectives at this level require understanding a content area so well as to put it in other word or form. It involves translation such as interpreting mathematics ideas from verbal form into symbolic statement or vice versa. It also involves the ability to make generalization of a given material.

The third category in the taxonomy is the application. This level of thinking involves the ability to apply rules, method and principles in solving mathematics in a new situation without any problem. Objectives at this level includes the use of words such as solve, give reasons for, prove and predict.

The fourth level is the analysis. This refers to a level of mental process, analysis emphasizes the breakdown of materials into component parts. It involves the ability to compare and constrast. This level may also involve the ability to identify elements which have not been clearly defined and at the same time involves the determination of the relationship between the various components of a mathematical structure.

The fifth level of taxanomy is the synthesis. This involves putting together ideas in order to form a whole; Obuche (1982). For example, students should be able to explain a mathematical word problem and put together all ideas he derives from the problem in order to solve it.

The last level according to Bloom is the evaluation stage of the cognitive domain. This according to measurement experts is the highest level of thinking. It involves making judgement. It implies using certain criteria as well as standards in appraising the extent to which particular outcomes are accurate.

However, Avital and Shettleworth(1968) felt that although Bloom's taxonomy is put in general terms so that it is not tied to a particular subject-matter, one could see that very few examples included in the terms are from mathematics. As a result of that, Avital and Shettleworth (1968) adopted and modified the major categories of Bloom's taxonomy to describe a model that explain precisely the levels of mathematic performance. Therefore, according to Avital and Shettleworth (1968) there would be three levels of thinking in mathematics which correspond to first five categories of Blooms educational objectives. These are:-

1. Thinking process, recognition and recall
2. Algorithmic thinking and generalization
3. Open search.

They evolve their model in such a way that thinking process recognition and recall take care of the first level of taxanomic level (Knowledge). The algorithmic thinking and generalization take care of comprehension and application and finally open search takes care of analysis and synthesis.

Another person again who evolved his own model was Wilson (1971). His mode of evaluating achievement test in mathematics contains four levels of behaviour in the cognitive domain. They are as follows:

(a) Computation items which are designed to test students knowledge of specific facts and terminology.

(b) Comprehension items are in this regard is designed to test students knowledge of rules, principles and generalization. Items designed here also test students ability to transform problems from one form to another.

(c) Analysis items according to Wilson tests students ability to construct his ability to solve *unconstructed problem to discover relationship.*

It should be of interest to note that some of these levels are within the scope of this research. This is because the of academic achievement and the content does warrant the use of the levels outlined above.

2.6 Evaluation Procedures.

The researcher feels that discussion on evaluation will not be complete if no mention is made of the formative, summative, criterion referenced and nonrm-referenced evaluation.

2.6.1 Formative Evaluation

Formative evaluation is used to monitor learning progress during instruction. Its major purpose is to provide continuous feedback to both pupils and teachers concerning

learning success and failures; Gronlund (1981). Perhaps the main purpose of formative evaluation is to determine mastery of a given learning task and pinpoint the part of the task not mastered; Adam (1966). The purpose is not to grade or certify the learner, it is to help the learner and the teacher focus upon the particular learning necessary for movement towards mastery. Gerberich (1964) noted that feedback to pupils provides reinforcement of successful learning and identify the specific learning error that are in need of correction. Gronlund (1981) said feedback to the teacher provides information for modifying instruction. Formative evaluation therefore, depends heavily on specially prepared test for each segment of instruction. It should be noted here that tests used for formative evaluation are most frequently teacher made.

We regard formative evaluation as useful not only for curriculum but also for instruction and student learning. Martuza (1977) contended that formative evaluation is for the use of systematic evaluation in the process of curriculum construction, teaching and learning for the purpose of improving classroom instruction. Since formative evaluation takes place during the formation stage, every effort should be made to use it to improve the learning process in the mathematic classroom.

2.6.2 Summative Evaluation

Unlike formative evaluation, summative evaluation typically comes at the end of the course of instruction. It is designed to determine the extent to which the instructional objectives have been achieved. It is also used primarily for assigning course grades or clarifying pupils' Mastery of an intended outcomes. Thorndike and Hagan (1969) pointed out that perhaps the essential characteristics of summative evaluation is that of making valued judgement about pupils progress at the end of any course.

The technique used in summative evaluation is determined by the overall instructional objectives and it typically include all forms of end of course examinations. Stanley (1972) said although the main purpose of Summative evaluation is grading or certification of pupils achievement, it also provides information for judging the appropriateness of the course objectives. We can say therefore that summative evaluation is directed towards a much more general assessment.

Having examined the summative and formative evaluation the researcher regarded his evaluation procedure as a summative type. This was because the test contained items that are geared towards assessing the overall mathematical achievement at the end of primary 3. Again it aimed at providing feedback, and reinforcement of successful learning at the end of a school year.

2.6.3 Norm And Criterion Referenced Measurement

The distinction between the two types of scores seem clear enough as pointed by Thorndike and Hagan (1977). Wood (1979) said if we interprete a score of an individual by compering his score to those of other individuals (called a norm group) this would be norm referencing. Marshall and Hales said if we interprete a person performance by compering it to some specified behavioural criterion of proficiency, this could be criterion referencing. To polarize the distinction we could say that the focus of normative score is on how many she/his peers do not perform (Score as well as he does, the focus of criterion referenced score according to Gronlund (1981) is on what is that Shehu can do. Of course, we often do interprete a single test score both ways. In norm referencing we might make a statement that Adam did better than 80 percent of the students in a test on addition to whole numbers. In criterion-referenced we might say Adamu got 70 percent of the items correct in a test of whole numbers.

Ebel (1979) pointed out that to be meaningful any test must be related to test content as well as the scores of other examinees. However, Millman (1980) agreed and correctly stated that any test will sample the content of some specified domain and that is always an explicit behavioural element.

In terms of uses Gaslin (1967) said that the principal uses of criterion-referenced measuremtnt has been in "Mastery test". A mastery test is a particular type of criterion-referenced test in which a person is tested as having achieved (mastered) the objective satisfactorily or he has not. Criterion-referenced in general could also apply to degrees of performance.

Clark (1968) said employing the individually preserved instruction or mastery model is not the only use for criterion-reference measures. In order to determine whether specific instructional treatments has been successful, it is necessary to have data about the outcomes on the specific objectives the programme was designed to teach.

Wood (1979) observed that criterion-referenced measures offer certain benefits for instructional decision making within the classroom setting. Turning to the uses of norm-referenced measurement. Many educators would agree with Gronlund's statement which says "In measuring the extent to which pupils are achieving their course objectives, we have no absolute standard by which to determine their progress. A pupil's achievement can be regarded as high or low by comparing it with the achievement of other pupils".

Looking at what norm-referenced is all about, the procedure adopted by the researcher could be termed as a norm-referenced one. This was true because of its emphasis on determining the extent to which the pupils had achieved a course objective at the end of primary 3 as compared to their classmates. Thus to say a pupils achievement can be regarded as high or low by comparing it with the achievement of other pupils. It can also spot out the area of strengths and weaknesses of the pupils.

2.7 Evaluation Instruments

By evaluation instrument we imply the use of various techniques in evaluating students progress. Evaluation techniques can be said to be using three main instruments as pointed by Gronlund (1981). These are ; (i) the measurement data which give a quantitative description of students performances, (ii) the non-measurement data which gives the qualitative description of the students work, and (iii) value judgement which is the worth of the result so obtained. Nevertheless, as pointed by Tuckman (1978) measurement is only restricted to quantitative explanation of the students performance in an achievement test. Thus to say, it answers only a very narrow question, How much a student gets in a test? For example Shehu scores 85% in arithmetic test at the end of

the term. It can be seen that it neither includes qualitative (non-measurement) nor does it employ judgement concerning the worth or value of the result obtained.

It is interesting to note that, when we interprete students performance, we just have to assign some values or worth on it in terms of other students in his class. This really shows that we are no longer restricting ourselves to quantitative description alone.

Generally speaking the most widely and frequently used measurements by the teachers and educators are tests. This is not far from the fact that well constructed tests give a good measure. Testing is the main technique used in this research.

2.8 Types Of Tests

In general terms, a test is an instrument designed for the measurement and evaluation of any knowledge, quality or ability. It may measure degree or amount of mental ability. Hills (1981) observed that test may be made up of items of similar difficulty and items arranged in increasing order of difficulty. Ebel (1979) defined testing as "The exposure of a client to any given device whether group or individual, essentially for the purpose of obtaining a quantitative characterization on one or more traits on that client".

He elaborates by suggesting "testing is the controlled observation of the behaviour of an individual to whom stimuli of known characteristics are applied in the known manner". Again Layman (1963) defined testing as "The evaluation of individuals or individuals performances on continua representing definable psychological traits and functions".

A sharp distinction should be made between the terms testing and assessment. The concept of assessment is distinctly different from that of testing. Kalejaiye (1982) said assessment involves obtaining a view of the individual's personality. It is therefore, agreed that the purpose of assessment is to put test scores in their proper context.

The objective test is perhaps, the most commonly used in our school system. The types that fall under this includes multiple choice, true-false, matching and short answers.

In Multiple Choice items, the examiner construct test items in which the candidates are required to select the correct answer from among a certain number of choices, say four or five. This type of objective test is widely applicable to various task performance. The only major disadvantage is that it is more difficult to construct than any other types. However, its main advantages are the objectivity in scoring and high reliability. Glass et al (1971) said multiple choice

test is expected to show much more reliability than most other types. However, critics of multiple choice have argued that it doesn't allow students to put in their point in a logical manner, they are only asked to select from alternatives which often give more room to guessing.

The true-false items require the candidate to either accept or reject a proposition. They are best suited for assessing knowledge that allow for true-false response. For instance "any quadritaterul is a polygon". One would see here that the candidate has 50 percent chance of getting the answer right or guessing. Thorndike and Hagan (1977) observed that the format yields less accurate information about a student per item than do other forms of items. Another disadvantage of this type is that many more items are needed in order to provide an equally precise of appraisal of each candidates competence. Hammil (1971) noted that there should be a relatively enough question and enough time for all pupils to read and respond to all of them.

Some common variations are Yes - No, right-wrong cluster variety and correction variety. One of the major disadvantages of true-false response is the use of negative statements and particularly double negatives. Wason (1961), working with adults and Zern (1967) working with children both have shown

that time required to answer negatively phrased item is longer than positively phrased item and that more errors are made to negatively phrased items.

The characteristics that distinguish the matching items from the ordinary multiple choice item is that in matching there are several problems whose answers must be drawn from a single list of possible answers. The matching item has most frequently been used to measure factual information such as meaning of terms, dates of events, the achievement of people etc.

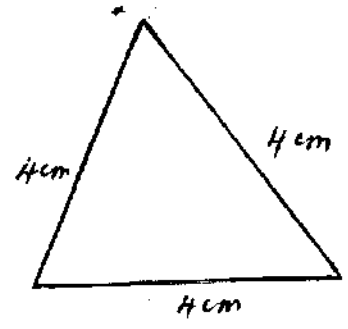
Gerberich (1962) said:-

> ...If all items on both sides are properly used in the matching process, the exercise has no distractors and is said to be balanced if, however, one list is longer than the other and indicates terms that are not involved in a parttern of correct answers, the exercise has distractors and is said to be unbalanced. p.236.

Gronlund (1981) said that items in the column for which a match is saught are called premises and the items in the column for the selection is made is called responses.

The next type of objective test item is the short-answer items. The short-answer item and the completion item are both supply-type test, items can be answered by a word, phrase

number or symbol. In this case the candidate is only required to provide short answer. For example, in the number 612, what value does the 6 represent? 600 or in the triangle below, what is the number of degrees in each angle (60°).



The short answer type of test is suitable for measuring a wide variety of relatively simple learning outcomes. The most important advantage of the short answer item arises out the fact that the students must supply the answer. Tukman (1978) observed that this reduced the possibility that students will get correct answer by guessing. They must either recall the information requested or make necessary computation to solve the problem presented to them.

Aptitude tests according to Gronlund (1981) are designed to predict future performance in some activity. In other words Aptitude test measure students ability to learn new tasks. Thorndike and Hagan (1977) viewed aptitude tests as the ones that largely possess in common the characteristics of testing the individuals potentialities in terms of specific abilities.

The most well known aptitude test is the intelligence test which is designed to measure learning ability. Merwin (1965) defined intelligence tests as the test that measure what is perhaps most simply and most commonly described as ability to learn. These tests are specially designed to measure a brand and overall intellectual functioning of a person. These types of tests differ from the aptitude tests which indicates the power to perform a certain task; Stanley and Hopkins (1972). That is to say the ability test refer to a current state and aptitude to a future state.

Wilson (1971) also considers achievement tests as measuring learning experience in a relatively restricted learning situation. The main interest here is to find out what has been learnt as a result of a particular learning opportunities offered. Achievement tests can take several formd. They can be paper- and pencil test or performance tests. Psychologists also saw that achievement tests can be on the basis of content which measure knowledge of some subject areas. In each of the subject area the achievement test can serve as a survey test which gives an estimate of a subject's overall level of performance in a content area. It can also serve as a diagnostic measure which in constrast to survey, attempts to assess the relative weakness and strength in important component skills. The diagnostic tests help to determine the area where an individual needs remedial work.

2.9 Planning The Test

Teachers planning their own objectives are found to face many problems. Gronlund (1981) pointed out that to remedy the situation the teachers must some how come to two decisions-what type of content to include in the test and what types of students behaviours or instructional outcomes, to measure Wallace (1978) observed that teachers must then effect a balance between test content and types of students behaviour to be measured. Jahun (1985) observed that teachers must then effect a balance between test content and types of students behaviour to be measured. Jahun (1985) recommended that a good test is the one that set exfalicit specification that include:-

1. Form of the test items to be used.
2. Number of items of each form.
3. Kinds of tasks the item will represent.
4. Number of tasks in each kind.
5. Areas of content to be covered.
6. Number of items in each area.
7. Level and distribution of item difficulty.

After the outline has been spelt out the test constructor will proceed to form the table of specification otherwise known as a blue-print. Preparing the two-way-grid in which the constructor would relate topics with objectives already set

ahead in the curriculum. This is considered to be the final step in the building of table of specification that relates to instructional objectives of the course content, and specifics the nature of the test sample.

2.10 Writing The Items

Items are written in due regard to the specification outline, thus to say that careful consideration should be given to each column put in the table of specification.

Measurement experts according to Gerberich et al (1962) suggested the use of more than one type of objective item in a test but on the other hand not to use too many variations. Fred (1967) suggested that:- The test maker will find its advantageous, to write each item or each test on a filing card or ship.

Lehman (1975) observed that items writing can be most effective when certain general principles and other details of constructions are observed.

2.11 Pretesting The Items

When a test constructor has finished writing his appropriate item and editing and revising, the items are arranged into test form. These should be administered as a preliminary stage to a sample of students who are similar to

the population that will take the final form of the test. At this juncture the test writer is expected to compute all the relevant statistics associated with the test. The aim of which is to eliminate some of the bad items and retain only the good ones. At this stage however, the constructor is expected to provide all the necessary instructions that will make the candidates taking the examination feel comfortable otherwise some extraneous variables may come into play to disrupt the examination.

2.12 Validation Process

Tuckman (1975) pointed out that in order to determine the validities of an individual item for future use, the method generally known as item counting is of great value. Ebel (1977) stated that:- "One of simple item counting methods is based on a devision of the class into groups of above average and below-average performance on the test, with half of the class in each group".

2.12.1 Preparing The Final Form

After the administration of the pretest and computing all the necessary satistical data, the selected items are then administered to a large population. The item prepered for administration must be well balanced to meet several conditions in order to produce tests with the desired

characteristics. They should definately represent the various skills as enumerated in the test plan. It should be noted that the number of items in each test set must be appropriate to the time limit, the item difficulty level and discriminating power (item analysis) must be determined.

2.12.2 Difficulty Index

In terms of objecgives test, difficulty takes on a sharper definition, and may be defined as a percentage of examinees who get the item right. For example if a particular item on a test is answered correctly by 40 percent of all the students who take the test, we may say the item has a 40 percent difficulty.

Usually as Tuckman (1978) pointed it, we talk about difficulty of a test in terms of average difficulty. For example, suppose we made up a test of 50 items and gave this test to a group of students. Assume that when we scored the test papers we found that the average score for all the students who took the test was 25 item right. Ebel (1979) said since the average score for the group of the students was 25 out of 50 we would then say that the test has an average difficulty of 50%. Mathematically difficulty index is calculated as follows:-

$$P = \frac{R}{T} \times 100$$

Where P = Percentage of student's who got the item right.

R = The number of students who got the item right.

T = Total number of students who attempted the item.

It should be noted that for both norm or criterion referenced tests, one useful index of an item's quality is its difficulty. Gronlund (1981) said that P values ranges from 0 to 1.0 with higher P values indicates item that more examinees answer correctly. For example an item with P value of 98 would be one which most examinees answered correctly. Similarly an item with value of 0.15 would be one which most examinees missed.

However, a note of caution should be registered at this point because according to Tuckman (1978) measurement people sometimes err by referring to items with high P values say 0.80 and above as "easy" items while items with low P values say 0.20 and below are described as difficult items. These ascertion may or may not be accurate.

2.12.3 Discrimination Power

This is an index which is used to show how an item differentiate between the more and less able examinees. Various indices have been used but a simple index which is practical for classroom teachers is the difference in the percentage of passing the item among those in the top and the bottom quarter in the total score on the test or on some other measures of the traits appraised by the test. The item discriminiting index is then computed by subtracting the number of candidates in the lower group who got the item right (refered to as R_1) from the number of candidate, in the upper group who got the item right (refered to as R_u) and dividing by one half of the total number of candidates included in the item analysis (refered to as $\frac{1}{2}T$); (Gerberich 1964). Thus the formula for computing discriminating power is as follows:-

$$D = \frac{R_u - R_i}{\frac{1}{2}T}$$

Mehrens and Lehman (1975) noted that for norm-referenced test one of the most powerful indication of an items quality is the item discrimination power. An item discrimative index typically tells us how frequently an item is answered correctly by those who perform well on the total test. In principle, an item discriminative index reflects the relationship between examinees responses on the total test and their responses on the particular test item.

In essence, a possitively discriminating item indicates that an item is answered correctly more often by those who score well on the total test than those who score poorly on the total test. A negatively descriminating item is answered correctly more often by those who score poorly on the total test than those who score well.

2.12.4 A Look At The Psychometric Qualities

A general concess from the measurement experts is that the determination of the psychometric qualititics of any test should be a continual process. Rummers (1960) observed that even though some variables such as time limits, difficulty level, distribution of scores etc can be computed from the pretest data students responses to the final test may differ from their responses to the pretests. This implies that analysis of the completed forms of the tests are needed. Empirical improvement of the test item is therefore sought for for in this direction.

2.12.5 Validity

When one is selecting or constructing an evaluation instrument, the most important question to ask is, "To what extent will the result serve the particular uses for which they intended"?; Gronlund (1981). Thus to say that validity

refers to the extent to which the results of an evaluation procedure serve the particular uses for which they are intended; Wilson (1971). If the results are to be used to describe pupil achievement we should like them to represent the specific achievement we wish to describe and to represent nothing eles; Thorndike and Hagan (1977). In his own contribution about validity Tuckman (1975) commented that validity is always specific to some particular use. For example the result in an arithmetic test may have a high degree of validity for indicating arithmetical reasoning, a moderate degree of validity for predicting success in future mathematics courses. There are four types of validity that are of interest to the measurement experts. They are as follows:-

(a) Predictive validity,

(b) Concurrent validity,

(c) Content validity,

(d) Construct validity.

Popham (1981) contented that the first two of these may be considered together as criterion-oriented validation procedures. The pattern of criterion-oriented study is familiar. The investigator here is primarily interested in some criterion which he wishes to predict. He administeres the test, obtain on independent criterion measure on some

subjects and compute a correlation. Brown (1979) said that:- If the criterion is obtained sometime after the test is given, he is studying predictive validity. If the test score and criterion score are determined at essentially the same time, he is studying concurrent validity.

Concurrent validity is studied when one test is proposed as a subtitute of other test.

Content validity is established by showing that the test items are sample of the universe in which the investigator is interested. In a study like this one content validity is very essential because of the level of the subjects taking the examination.

Construct validity as pointed by Scott in (1967) is involved whenever a test is to be interpreted as a measure of some attribute or quality which is not operationaly defined. Annastasi (1959) commented that construct validity is important at any time for every sort of psychological tests aptitude, achievement, interest and so on. Thurstone (1969) statement said:- "In the field of intellegence test, it is used to be common to define validity as a correlation between a test score and some outside criterion".

2.12.6 Reliability

Considering a teacher built-test to measure achievement or learning, the teacher is really interested in how much the student has learned, that is the learning score. Marshall (1971) reported that to determine that score, the teacher must construct a measuring instrument called a test. However, balance the test result reflects not only how much the student has learned, but also how accurate a measuring instrument is, the score for each student is less than perfect measure of how much he or she has learned.

Stanley (1972) commented that:- The extent to which a particular test give a consistent and accurate representation of the property being measured by the test is a function of the test reliability.

Reliability indicates the degree to which the test is consistent in measuring whatever it does measure-degree to which the test measures the same thing time and time after item; Cronbach (1970).

Ebel (1977) pointed out that:- In order a test to be valid, it must be reliable, unreliable test cannot possibly be valid, merely because a test is reliable, however, does not guarantee its validity.

In order to have an accurate reliability of one's test the following have to be taken into consideration.

2.12.6.1 Stability

Stability estimates of reliability, are based on the consistency of a test measurement over time. Wood (1971) observed that the most common way of securing a test stability is one should administer a test to a group of examinees, wait for a reasonable interval than readminister the test to the same examinees. By correlating examinees scores on the two tests we secure a reliability coefficient.

It should be noted however, that the length of time between the two testing occasions is of course, an issue of significance Brown (1983) said if we wait only for a few minutes to readminister the test, the examinees second performance is likely to be influenced by the initial test performance and vice-versa.

2.12.6.2 Equivalence

The second form of reliability is associated with the use of two or more forms of the same test, since it is generally hoped that the two tests or more forms will yield consistent information regarding examinees.

Ebel (1979) said the most straight forward manner of securing equivalence information is to administer two forms of the same test, then correlate examinees scores on the two forms. The resulting coefficient is refered to as the equivalence coefficient of reliability.

The coefficient of equivalence between scores on parallel forms of the test, administered in close succession can be calculated according to Brown (1979) by using the formular.

2.13 Studies Outside Nigeria

The researcher has made several efforts to find out some studies that have been conducted outside Nigeria which have some connections with the present study. It is really very difficult to get an identical studies, however, the researcher was able to get some that are somewhat similar to the research topic. It should be noted here that most of the research in measurement and evaluation are carried on in the United State of America with few ones in other part of the world.

Costell (1975) reported that in a study he conducted on how socio-economic factor affects the school achievement test. He found that there is a large variation of cultural difference among cultural groups. At one time these variations used to be incorporated in terms of the inferiority of one cultural group compared to another.

His result showed that young children's achievement in Mathematics learning was observed to be related to age and intelligence. Socio-economic level was also found to be related to achievement, though not so strongly.

George (1971) prepared college board schorlastic aptitude test in secondary school in England. A form of schorlastic aptitude test as a predictor of academic achievement was administered to a sample of 1,008 students in four boys schools and four girls schools in England in March (1971). At each school either all of the fifth-form or all of the sixth-form students were tested and they found the following results.

1. The British students performed very well in SAT and found that the obtained scores were reliable and proved valid predictor of grades on essay examination.
2. Each SAT-verbal item type showd about the same correlation with total score as obtained in the United States.
3. SAT-verbal and SAT-Mathematical scores proved to be quite reliable for the British Students. Unsing Kunder-Richardson reliabilities fomula 20 it was found that the reliabilities were .87 verbal and

.89 mathematical for Southern group, .88 verbal and .89 Mathematical for Northern group.

It shows that the mathematical score for the British fifth-form group are substantially higher than those of the U.S.A. group who successfully completed their first year of college. The relatively higher mathematics scores of British students might not be unconnected with greater emphasis on mathematics training they give to their students.

Stanley et al (1967) in a similar study took an original pool of 75 items out of which the investigators agreed that 59 of the items had content validity. The 59 items scale was administered to 35 undergraduates in introductory psychology. The reliability of the test using Kader-Richardson 20 formula was .85 and 36 of the items had point basireal correlation with a total score that was significant beyond the 0.05 level.

Alan (1971) prepared an achievement test reliability as a function of pupils-response procedures. The study was designed in part to assess the differential effect of two pupils response procedures (answering directly in the booklet versus on a separate answer folder) on Matripolitan Achievement Test (MAT) scores of grade 3 and 4 pupils.

Over 4000 pupils from nine school systems took the matripolition test half responding in their booklet and half using answer sheet.

Although the separate answer sheet group received lower scores than their counterpart responding in the test booklets, the results indicate that the reliability of scores is not significantly altered when pupils respond to achievement test items on separate answer sheet rather than directly in their booklets. Contrary to the finding of Muller et al (1966), it does not appear that the use by pupils of response media separate from those test booklets adversely affects the reliability of scores derived from the various tests of the Matripolitan Achievement Tests.

2.14 Studies Within Nigeria

As pointed out by Ikeotuonye (1982) Fomal testing programme in education started in Nigeria with the establishment of Western Education. There is no doubt to claim that it was the missionary, through their mission schools brought the idea of examinations in Nigeria. As the number of schools began to rise in West Africa, the need for a unified body to cater for examination was sought for. Jahun (1985) reported that Ohuche and Akegu (1977) contended that the ordinance to the establishment of this body was passed in Accra Ghana in 1953.

Obioma (1985) prepared a 60-item diagnostic mathematics test (DAMAT) for the junior secondary schools in Anambra, Cross River and Benue States. He found that the item retention rages (IRR) which defines the various item facilities (PACS) were pre-specified to guide the process of item modification.

The intercorrelation among the DAMAT scales and complexity demension were generally low, yet the internal consistancies of the DAMAT comparatively was high (α = 0.77). The DAMAT scales and its complexity demension had consistancies in the range of 0.37 - 0.5.

Taken together, as pointed out by Obioma, DAMAT psychometric measures on inter-correlation and internal consistancy show that DAMAT is an effective diagnostic tool. This is so since low inter-correlation accompanied by high consistancies are evidence of high effectiveness diagnostic tools.

A validation studies of the Differential Aptitude Tests (DAT) was carried out in Kaduna State by Ikeotuonye (1982) to determine the applicability of DAT in Kaduna State for the purpose of educational and vocational guidance. The test were given to a stratified random sample of 400 secondary students whose ages ranged from 13-23 years. The reliability coefficient observed from the study ranged from 0.34 to 0.94 and the standard errors of measurement ranged from 0.33 to 2.86.

Jahun (1985) in an attempt to validate and standardized algebraic tests for Junior Secondary Schools in Kaduna and Kano States constructed five parallel tests. He administered these items to 349 students in form three and 659 students beginning form four.

On the basis of item analysis, he selected to sets of parallel tests, and administered them to 3,081 form three students in 14 schools randomly selected in Kano State. From his study he established three different far norms. These are norms of the total t sample, sex norms and institution type norms for each test set. Altogether he identified 10 different norms. He found that the reliability coefficients ranged from .78 to .89 and eight of the ten different norms falls within the .80's. No validity index was ever reported.

2.15 Sampling Techniques

Whenever a researcher is to embark on a research he must decide on the type of sampling procedures he would adopt for his research. Several methods of the techniques have been devised by experts. They include sample random sampling, stratified sampling, systematic and cluster sampling.

2.15.1 Random Sampling

In random sampling as pointed by Best (1970) every element in the population has an equal chance of being selected. This implies that no element of the population has been omitted deliberately by chance. Tuckman (1978) said an important distinition must be pointed out between random samples and representative samples. A simple random sample is one drawn in such a way that each member of the population has an equal chance of being selected. On the other hand, in a representative sample cases are selected in such a way that characteristic of the sample are similar in an important respects to the characteristics of the population.

2.15.2 Stratified Sampling

Stratified random sampling is the process of dividing a population into strate or sub-population; Cohen et al (1981). Also commenting on stratified sampling Fox (1969) said after a population has been divided into strata, samples are selected randomly but independently from each stratum. Indagi (1981) observed that with stratified samples, the attempt is to maximize differences between strata and minimize differences within each stratum.

2.15.3 Cluster Sampling

Sax (1968) said in the case of cluster, the researcher samples areas of cluster first, and then samples individual elements within the cluster. Osuale (1982) however, pointed out that cluster sampling usually produces a large sampling error then a simple random sampling of the same size. This is because each cluster may compose units that are like each other and this reduces representativeness of the sample.

Having examined each sampling technique critically, the researcher therefore found stratified random sampling to be the most useful to be employed in a research of this nature. This is because the population is composed of various characteristics which need to be taken care of seriously.

2.16 Summary Of The Literature Survey

It is apparent from the foregoing review that psychologists and mathematics educators have all agree that concrete operational stage is a very vital stage of the child's development. The critical workd of both Piaget and Brunner, triggered off radical changes in Mathematics pedagogy resulting in the evaluation of never strategies to mathematics teaching and learning and all of which emphasize that children at this stage of development needs to be properly handled. Several concepts, such as addition, subtraction, multiplication,

division and fractions can hardly be learnt effectively at this stage without a considerable amount of teaching strategies.

In the literature surveyed, several issues on the concept of evaluation at elementary level remained unexplained. For example, there were no statement of the relative importance of test construction in mathematics at this level. Neither was there a need for a uniform validated mathematics test in the primary schools.

The review showed that there is a need for teachers to take a serious look at evaluation model when constructing their mathematics tests. The survey reveal that only teachers of mathematics outside Africa do take such into consideration while teachers in Black Africa don't.

A critical look at the evaluation procedures reveals that most of our test reporting tends to favour only criterion-referenced at the expense of norm-referenced.

From the survey, measurement experts highlighted the importance of item analysis and psychometric qualities of any given test. For example, they sighted many advantages of these characteristics such as Discriminative Index, Difficulty levels, Reliability, and Validity.

This study is designed to construct and validate a comprehensive test items in Primary 3 Mathematics. Investigations from the research works conducted within Nigeria shows that only a few people have worked in this area of study. Among the few that were done, all were secondary level and none from primary school level.

Chapter 3

RESEARCH METHODOLOGY

3.1 Introdudtion

The research was intended to construct and validate test items from the contemporary primary 3 mathematics curriculum. In the first place, the researcher started the research by going on a familiarization tour to some Primary schools within Zaria Local Government Area. The objective of the tour was to find out whether all the primary schools visited are operating on a similar mathematics curriculum. The results of the tour revealed that all the schools visited are operating on the same curriculum that is, the New National Primary Education Mathematics Curriculum.

Having seen that the schools are operating the same curriculum, the researcher then embarked on the construction of the table of specification otherwise known as blue print. Thus a total of 160 questions that were believed to be within the scope of the primary 3 pupils were prepared and properly validated by experts. The methodology used in this research involved two stages, the first stage is called the first try out and the second is the final try out.

The Population Under Study

It should be noted here that the population of this study are the primary 3 pupils in Zaria Local Government Area. With the creation of the New Kaduna State, Zaria Local Government Area becomes the largest among the seven local government areas in the State.

In order to accomplish the objective of the study the entire population of the primary pupils in class 3 in the Local Government totalling about 19,037 from 147 primary schools in the local government were considered. It was observed that the Local Government Area consists of five districts; namely Sabon Gari, Giwa, Soba, Igabi and Zaria City and its environs.

Sample for the First Try Out

For the proper execution of the first try-out only subjects from Zaria matripolitan area comprising Samaru, Sabon Gari, Musasa, Zaria City and Tudun Wada formed the sample for the first try-out. Among the schools in the matripolitan area, five primary schools were considered to take the first try-out.

Although there are various techniques of sampling in order to obtain an unbiased sample, for the purpose of this study, the researcher used stratified random sample. This

was because it would give a fair representation of all the various types of schools with their different cultural and academic background. Thus the researcher stratify the metripolitan into two strata namely Zaria City and its envorons and Sabon Gari District.

Investigation carried out by the researcher revealed that Zaria city and its environs has a total of 50 primary schools while Sabon Gari district has a total of 36 primary schools.

The following computation was carried out by the researcher which gives the number of schools to be considered from each stratum.

Number of sample from each stratum $\frac{n_1 \times H}{N}$

Where

n_1 = no of schools in a strata.

N = no of schools in the metripolitan

H = no of schools used as a sample.

The result of the computation shows that.

1. 3 schools were selected from Zaria City area
2. 2 schools were selected from Sabon Gari District.

The schools were then selected at random by hat-draw method.

Table 3.1: The schools selected for the First Try-out in Zaria Matripolitan.

S/NO	Schools	District	No of Pupils in the class	No of those who participated
1.	Hanya Children School Zaria	Zaria	64	64
2.	L.E.A. Bomo	Sabon Gari	47	47
3.	L.E.A. Dakache	Zaria	90	60
4.	L.E.A. Unguwar Limun Zaria City	Zaria	160	100
5.	A.B.U. Staff School	Sabon Gari	240	94

In the first three schools in table 3.1, the researcher found that they have only two arms of class 3 so the first try-out was taken by the pupils in the two classes found from each school. Wheares in the remaining two schools have more than three arms of class 3, the researcher randomly picked three arms which gives the total number of pupils that perticipated in the examination from each school.

However, the result obtained from the first try-out which could be seen later in this chapter showed that the random sampling employed by the researcher was suitable for this kind of study. Hence, the researcher feels that the same kind of sampling technique be maintained for the final try-out.

Sample For The Final Try-Out

The subjects for this study were the primary 3 pupils from Zaria Local Government Area. The subjects from Zaria Metripolitan were used for the first try-out. It should be noted that the purpose of sampling is to achieve a fair representation of all the characteristics of the parameters of the parent population.

The various characteristics of the parent population which include urban and rural areas, public and private schools, facilities and lack of qualified teachers in most of the schools in the rural areas as well as in some parts of the urban areas, it therefore became imperative for the researcher to categorize the population in two ways.

(a) Districts

(b) Public and Private Schools.

In order to ensure that each of the characteristics mentioned above was well represented a stratified random sampling as earlier mentioned was employed. With this process, the population was stratified as follows:-

1. The urban public schools
2. The urban private schools
3. The rural public schools
4. The rural private schools.

Having stratified the population, it was found that 55 public schools fall under urban public primary schools, 14 private schools fall under urban private schools, 80 public primary schools fall under rural public school while none private school fall under Rural Private Schools. The same arithmetic computation as mentioned in paragraph 3.3 was used and came up with the following results:-

The seven urban public schools are the following:- L.E.A. Schools; Kofar Jatau, Gyellesu, Palladan, Jaji, Lemu Samaru, Dambo and Dogarawa. The eleven rural public schools are; L.E.A. Schools; Turawa, Soba, Yakawada, Farakwai, Awai and Gaden Gayan. The two private schools are the Nigerian Railway Staff School and Uphill Primary School Unguwar Kaya Zaria.

Item Distribution

There researcher first started the construction of the test items by thoroughly analyzing the primary 3 mathematics curriculum (See in Appendix A). A two-way grid was then developed to relate the content with the behaviour outcomes expected at the primary 3 mathematics course with the view to forming a model of the test plan.

After analyzing the curriculum, it was found that the content has four main topics. These are:-

1. Number and Numeration
2. Basic operation
3. Measurement
4. Descriptive Geometry.

It should be recalled that Jahun (1985) came up with a model of construction of table of specification to suit his own research. His model is different from that of Bloom's model discussed in chapter one because his own contains only the following levels.

1. Knowledge
2. Mathematical Skills
3. Problem solving.

Having considered the level of maturity and educational standard of the population of this study. Jahun's model was modified to give the following behaviour levels.

1. Knowledge
2. Computational skills.

This was because the children at this stage of development as pointed by Piaget (1957) are at their concrete operational stage. Piaget (1959) said what the Pupils can do in mathematics at this stage is to handle things which require them to recall,

define or at most engage in simple computation. This implies that any mathematical task at primary 3 should be limited to these cognitive domain of educational objectives.

After a critical look at the curriculum the content and behaviour levels were expanded to include a detailed breakdown of the content and expected behaviour outcomes. The content areas were therefore expanded as follows:-

Number and Numeration

1. Whole number
2. Fraction.

Basic Operations

1. Addition
2. Subtraction
3. Multiplication
4. Division.

Measurements

1. Money
2. Measuration
3. Time.

Descriptive Geometry.

1. Identification of various shapes
2. Counting of corners, faces no of edges and no of sides.

While the behaviour levels were expanded as:-

Knowledge	Computational Skills
1. Concepts	
2. Generalization	1. Computation
3. Rules	2. Routine problems.

This becomes the model which the researcher feels is in line with the nature of the course content and the level of academic standard of the population under the study.

After deciding upon the cognitive behaviours to use, table 3.3 was developed to guide the construction of the test items which would test the expected behaviour outcomes.

Table 3.2: Guideline for constructing the test Items for validation.

Content	Knowledge	Computational Skills	Total	Percentage
Number and Numeration	6	1	7	23%
Basic Operation	3	11	14	46%
Discriptive	2	0	2	6%
Geometry				
Total	15	15	30	100%
Percentage	50%	50%		

The proportion of each cell in table 3.3 was decided based on the relative importance attached, the time that would likely be spent on treating each topic and coverage given to the topic in the textbooks used in schools. The researcher was told by most of the teachers that in primary 3 more emphasis in terms of length and time is given to computational problems, hence largest allocation of items was given to Basic operations.

Table 3.3: Detailed Table of Specification for Test Validation for Primary 3 Mathematics Achievement Test.

CONTENT	BEHAVIOR					
	KNOWLEDGE			COMPUTATIONAL SKILLS		
	Concepts	Generalization	Rules	Computation	Routine Problems	TOTAL
Number and Numeration						
Whole number	1	1	2			4
Fractions	1	1		1		3
Basic Operations						
Addition			1	2	1	4
Subtraction			1	2	1	4
Multiplication			1	2	1	4
Division				2		2
Measurements						
Money				2	2	4
Measuration				1	1	2
Time	1					
Discriptive Geometry						
Identifications of various shapes	1					1
Counting of Corners faces etc	1					1
TOTAL	5	2	5	12	6	30

Item Construction

The books written for the New Curriculum under study were used to construct the test items. These books are the most widely used books in all the primary schools in Kaduna State. They include Macmillan primary schools mathematics by P.N. Lassa, at el. The oxford Arithmetic course by P.N. Lassa, C.N. Okosi and J.A.O. Sofolahan, A.U.P. Primary Mathematics by N.A. Adumanu at el. and Longman Junior Decimal Mathematics books 3 by N.V. Brindley which is not in our primary schools nowadays but contains some topics that are in line with the new curriculum.

The use of objective type of test was motivated by the fact that most pupils in primary schools particularly those in primary 3 would find it difficult to comprehend essay type of questions. Moreso most of the books written for this level of education contain questions that call for only objective type of test. Another reason for using objective type is that the scoring of the items could be very objective and faster using a carefully prepared scoring key.

After writing the items, a copy of each test was given to some primary school teachers of class 3 in the schools visited by the researcher and three lecturers in the Department of Education A.B.U. for content validation.

The teachers and lecturers both responded positively by going through the 160 items constructed in the blue-print and passed their comments and suggestions. Their comments and suggestions were incoporated and this helped the researcher to come up with the three sets of tests for the first try-out. These three sets of tests were given to some teachers in staff school A.B.U. and L.E.A. Samaru for their criticism. The teachers accepted that the items were equally weighted in terms of content and difficulty level. When these priliminary works on items constructions were finished, the tests were administered for the first try-out.

First Try-Out.

The three sets of tests were administered to primary 3 pupils. It is worthwhile to note that the tests were administered under the close supervision of the researcher with some assistance from the teachers in all the schools. The two private schools namely A.B.U. Staff School and Wanya Children School took their examination on the 12th November, 1987, while the remaining public schools took their own on the 13th November, 1987. The idea behind this arrangement was to prevent any possibility of leakage. To strengthen this point further, the researcher collected both questions papers and answer

scripts from the pupils. The scoring of the tests was done by the researcher and each test paper and was marked out of 30 and the result for each pupil was recorded. A total of three hundred and sixty five (365) pupils paticipated in the first try-out. The breakdown shows that 121 pupils took test I, 115 took test II, while 128 pupils went in for test III.

Statistical Analysis For The First Try-Out

The three test were subjected to statistical analysis in order to calculate the means and standard deviations of each test set. The analysis also involved the calculation of the descrimination power and difficulty level of each item in the three tests.

Table 3.4 Means and Standard Deviations of the Test Sets

(N = 365)

Test Set	Max Score	Mean	Std Deviation	No of Cases
1	30	11.93	9.65	121
2	30	12.55	9.74	116
3	30	12.27	10.74	128

Table 3.4 above shows the means and standard deviation of the three tests. Both the means and standard deviations of the three tests are almost the same.

In order to justify the sampling technique employed for the first try-out the means and standard deviations of the schools that participated were computed and presented in Table 3.5

Table 3.5 Means and Standard Deviations of the Schools for the first Try-Out.

Schools		Mean	STD Dev.	No of Cases
1.	Hanya Children School Zaria *	18.90	7.05	64
2.	L.E.A. Bomo	4.67	2.53	47
3.	L.E.A. Dakache	13.78	5.34	60
4.	L.E.A. Ung. Liman	2.86	4.36	100
5.	A.B.U. Staff School*	21.04	6.68	94.

The table shows that the means of the two private schools marked with astrick are higher than the means of local Education Authority types of schools. The A.B.U. Staff School has a mean of 21.04 out of 30. In School number 4, that is L.E.A. Unguwar Liman Zaria city, the mean performance is 2.89 and the standard deviation is even bigger than the mean. This means that zero was within I standard deviation below the mean since no one can score below zero, the distribution is therefore skewed to the right.

Looking at the performance once again one could see that the high performance of the private schools could be attributed to the fact they have better facilities and have more qualified teachers than their counterparts in the public schools. This statistics further confirmed the assumptions that the private schools perform better than the public schools. Another reason was indicated at the examination when some pupils in the public schools told the researcher that they were not familiar with this type of testing, while their counterparts in privates schools were very much familiar with the system and some were even complaining that some of the items were simple.

Table 3.6 Mean and Standard Deviation of the Sex

Sex	Max Score	Mean	STD Dev.	No. of Cases
Boys	30	12.39	9.60	200
Girls	30	12.08	9.44	165.

The table 3.6 shows that the means and standard deviations of both boys and girls are almost the same. Thus to say there is no significant difference between the performance of boys and that of girls at primary 3 level of education. However, for the researcher to obtain an acceptable judgement between the performance of the two groups a test of significant difference of means was computed

manually as shown below. Since the data samples are independent, the three population variances are not some, that is $\sigma_1 \neq \sigma_2$, hence the following test statistic formulate was applied.

$$t = \frac{X_1 - X_2}{\sqrt{S^2_{1/N_1} + S^2_{2/N_2}}}$$ Perguson (1981, P.178).

Where

X_1 = Means of the boys

X_2 = Mean of the girls

S_1^2 = Variance of the boys

S_2^2 = Variance of the girls

N_1 = No of boys

N_2 = No of girls.

From the computer print out the researcher found that the means and variance above were computed as follows:-

$X_1 = 12.38$ $X_2 = 12.08$

$S_1^2 = 90.58$ $S_2^2 = 89.17$

$N_1 = 200$ $N_2 = 165$.

The test of difference of means of 0.30 was obtained which is less than 1.64 for = 0.05 indicates that there is no significant difference between the boy's

performance and that of the girls. This result tells us that at primary 3 level boys and girls perform relatively the same. Perhaps the difference exists only at the higher classes of primary schools.

The second step of the analysis was the determination of the difficulty level and the discriminating power. The item difficulty level was calculated for each item using all the pupils who participated in the examination. It is of interest to note that for both norm and creterion-referenced tests, one useful index of an item quality is its difficulty. The most commonly employed item-difficulty index often refered to these days as P-value is calculated as follows:-

$$P = \frac{R}{T} \times 100$$

Where

P = Difficulty index

R = The number of candidates who got the right

T = The total number of candidates who tried the item.

Popham 1981.

For norm referenced tests, one of the most powerful indicator of an item's quality is the discrimination index. In brief, an item discrimination index typically tells us how frequently an item is answered correctly by those who

perform well on the total test compared with those who score low overall. Thus to say that the discriminating power of an item differentiates between pupils with high and low achievement. It is obtained by subtracting the number of pupils who got an item correct in lower group from the number of pupils who got the item correct in the upper group and dividing by the number of pupils in each of the two groups; Gronlund (1981).

For the purpose of this research the discriminating power was calculated using the candidates who took each set of the test and scored within the lower and upper 27% of the distribution. The following procedures as suggested by Popham (1981) were used:-

1. Order the test papers from high to law by total scose.
2. Divide the papers into high group and low group by taking the lower and the upper 27% of the papers.
3. Calculate a P - value for each of the high and low group ie the determination of the number of examinees in high group who answered the item correctly, then divide this number by the number of the examinees in the high group. This provides us with P_n. Repeat the process for the lower group to obtain P_v.

4. Subtract P_v from P_n to obtain each item's discriminating power (D). In essence, then

$$D = P_n - P_v$$

Popham (1981).

The difficulty level was calculated using all the candidates who took each test with the help of the computer. It should be pointed out here, that the computation of the difficulty level using the computer took the researcher well above a month. This is not far from the fact that the computer itself has of the numerous problems associated with it. However, the discriminating power was computed manually because the SPSS Package in Iyah Abubakar computer centre does not have a programme as of now for the computation of the discriminating power.

On the bases of the difficulty level two sets of tests would be selected for the final try-out. Looking at both the discriminating power and the difficulty level correctly, the researcher considered only items between .35 to .60 levels of the difficulties for each item from 1 to 30 for the three sets of tests, picking two at a time. The purpose of this procedure however, was to pick items that majority of the pupils would be able to answer since what was expected was to come up with test items an average child should be able to answer. Based on

what has been said above, the items that were marked "X" were the ones selected for the final try out.

Appendix C shows the analysis in which the difficulty level was computed by the computer and the discrimination power was computed manually.

The items marked "X" were the ones selected for the final try-out. These items gave out two sets of tests which were administered to 20 primary schools mentioned earlier. Since the research concerns only construction and validation of test items not standardization, the final administration took take place at the end second term.

However, a point of clarification had to be made here about the actual class that took the final try-out. It could be recalled that the first try-out was given in November, 1987 which coinceded with the time the present class four pupils were in the third term of class 3 last session. Since the test was designed for the pupils in primary 3, the final try-out was administered to the present primary 3 pupils when they reached the end of the second term. This was because by then, it was hoped that the pupils might cover 90% of the syllabus designed for the primary 3 pupils.

The Administration Of The Final Test

At the proporsal defence it was resolved that the following recommendations be implemented:-

(i) That the alternatives be increased from 3 to 4

(ii) The test items be translated into Hausa and a tape recorder be used to give instructions to the pupils. The decision to use Hausa instead of English came as a result of the fact that most of the pupils in the Local Government Area found it difficult to comprehend the English language used in the first try-out.

However, because of the cosmopolitant nature of the sample, the researcher in consultation with the supervisor had to deviced an effective way of administering the examination. Eventhough it was agreed that the test items be translated into Hausa it was discovered that there were some pupils who could not read Hausa very well. In order to cater for these sets of purpils, the following schools which have most of their pupils as non Hausa speakers took their examination in English with English tape recorder as a medium of instruction, see appendix D for the English items:-

(i) Nigerian Railway Staff School

(ii) L.E.A. Lemu Samaru Zaria.

(iii) L.E.A. Gyellesu Zaria

(iv) Uphill Primary School Unguwar Kaya Zaria

(v) L.E.A. Jaji.

The remaining schools took their examination in Hausa with Hausa recorder as a medium of instruction, see appendix E for Hausa items.

Procedures For Administering The Final Test

This section describes the step by step procedures for administering the final testing to primary 3 pupils in Mathematics. Seven hundred and seventy two (772) primary 3 pupils were the subjects that took the final testing. The nature of the study involves a journey to all the schools and time consuming in conducting the two tests it was therefore felt that 36 subjects per representative school were adequate.

To ensure that the subjects were truly representative of their schools all the primary 3 pupils in each school were stratified into three categories of good, average, and weak. The achievement categories were based on the pupils most recent first term examination. From each achievement category seven boys and five girls were randomly selected so that each school produced 21 boys and 15 girls. The method is to ensure that the achievement and sex of the pupils were equally represented. Lunsdaine

(1963) found that when dealing with a large population stratefied random assignment of individual was better than treatment of groups based on class. This is because stratified random sample has the advantage of reducing error variance and of rendering the comparisons among samples potentially more sensitive.

The administration of the two tests took place on different dates in all the districts in the local government in the first week of July, 1988. The post test in some selected schools took place two weeks after the administration of the first tests.

The Statistical Analysis Of The Final Test

The statistical analysis of the final test results involved the followings:-

(i) Calculation of means and standard deviations of the various variables involved in the study.

(ii) Calculation of Discrimination index and difficulty levels which gave the final item analysis.

(iii) The testing of the null-hypothesis

(iv) The estimation of reliability.

(v) The estimation of concurrent validity.

Means And Standards Deviations Of The Various Variables In The Study.

The students raw scores were used for the calculation of means and standard deviation. These were calculated so as to give the means which shows the strength of each variable and the standard deviations which show the spread of each variable from the mean.

Discrimination Index And Difficulty Levels

The following formula was used for the computation of the item difficulty level and the same procedures as earlier indicated in chapter 3 was used.

$$P = \frac{R}{T} \times 100$$

Statistical Design for The Data Analysis

To compare the performance of pupils in Mathematics' achievement test, the mean score, the standard deviations and the number of students in each sample were used for the following groups of variables in each test i.e. test I and test II.

(a) Pupils who took test I

(b) Pupils who took test II

(c) Pupils from urban schools

(d) Pupils from Rural Schools

(e) Boys from urban schools who took test I

(f) Boys from rural schools who took test I

(g) Girls from urban schools who took test I

(h) Girls from rural schools who took test I

(i) Boys from urban schools who took test II

(j) Boys from rural schools who took test II

(k) Girls from urban schools who took test II

(l) Girls from rural schools who took test II

For the comparison of group (a) and group (b) the total number of pupils in each group were used. The mean scores of the two groups X_{PI} and X_{PII} for group (a) and (b) respectively, were compared using the difference of means. Since the number of pupils in each group was fairly large the scores were assumed to be normally distributed and since only two cells were involved. Hence the statistic used to test the hypotheses was the t-statistics. The level of significance adopted to test the hypotheses was 0.05.

For the comparison of groups (c) and (d) the total number of pupils in each group were used. The mean scores of the two groups X_{Pu} and X_{Pr} for group (c) and (d) respectively, were compared using the difference of means. The same t-statistic was used for the testing

of the hypothesis at 0.05 level of significance, because, theoretically one would expect group (c) to perform better than group (d). The same statistical design was used for the testing of the remaining hypotheses.

Reliability Estimates

The raw scores of the pupils in five schools who took both forms of the tests were used for the estimation of the reliability coefficient. Hence Pearson Product moment (rho) formula was used.

$$r = \frac{N\Sigma XY - (\Sigma x)(\Sigma Y)}{\sqrt{\left[(N\Sigma x^2 - (\Sigma x)^2\right]\left[(N\Sigma Y^2 - (\Sigma y)^2\right]}}$$

Starndard Error Of Measurement

It should be recalled that for one to be really informed about the accuracy of test, one needs to know its standard error of measurement at different score levels. To this end therefore Popham (1981) defined standard error of measurement as an estimate of the standard deviation that would result for the score of an individual if it were possible to measure him a number of times.

The computer gave out the standard errors of measurement of each variable involved in the test and the result was tabulated in the next chapter.

Concurrent Validity

The concurrent validity was computed using the observed test results and the pupils recent mathematics test results conducted by their class teachers. For this, however, the statistical design used was Pearson Product moment (rho).

$$r = \frac{N\Sigma x\ Y - (\Sigma x)(\Sigma Y)}{\sqrt{\left[N\Sigma x^2 - (\Sigma x)^2\right]\left[N\Sigma Y^2 - (\Sigma Y)^2\right]}}$$

Summary

The methods used for obtaining samples for data collection were (i) sample random sampling and (ii) stratified random sampling. For the analysis of data in the first try-out analysis of variance was used to test the parallelness of the three sets of tests. Items analysis was done which resulted in coming out with two sets of tests that were used in the final administration. The t-statistic was used for the testing of the hypotheses

The Pearson Product moment (rho) was the statistic used to estamate both the reliability and content validity.

Chapter 4

DATA PRESENTATION AND ANALYSIS OF DATA

4.1 Introduction

In this chapter a summary of data collected was presented. The results of the data analysis and finding were presented. The first part of the analysis delt with the comparison of the two sets of tests to find out whether the two test are parallel. The second part of the analysis as shown in Appendix F delt with the item analysis of the test results. The third part delt with testing of the null hypotheses, while the last part delt with the estimation of the reliability and concurrent validity of the two tests.

4.2 Testing For Equivalence By The Use of Means and Variances Of The Two Test Taken By The Same Group.

For this analysis the means and standard deviation of the Two tests taken by the same group of students was computed as shown in table 4.1 below.

Table 4.1 Means and Standard Deviation of The Two Tests Taken by the Same Group.

Test Set	Max Score	Mean	Std Dev.	No of Cases
1	30	14.65	5.71	166
2	30	14.08	5.64	166

On inspection of table 4.1 above it shows that means of the two sets of tests are almost the same while the standard deviations seem to be very close, Karkani (1968) and Devis (1975) have indicated that one method of estasblishing the degree of equivalence of any two tests is to observe that the means and standard deviations of the tests scores are the same or at least not significantly differecnt statistically

In order to conclusively determine the equivalence of the two tests, the t-tests statistical technique was computed in respect of the means scores and standard deviations of the two tests taken by the same group.

Table 4.2 t- test of the Test Sets.

Test Set	Max Score	Means	Std Dev.	Df.	t-value
1	30	14.65	5.78	165	0.91
2	30	14.08	5.64	165	

From table 4.2 above it is evident that the calculated t-value of 0.91 shows that there is no significant difference between the means of the two tests. The F-Value of the test of significance

between variances is 1.025 as shown in table 4.2.1 below.

Table 4.2.1 F-Test for Standard Deviations

Test Set	No of Cases	Variance	F-Value
1	166	32.60	1.025
2.	166	31.81	

The value shows that there is no significant difference in the variability of the scores of the two tests at .05 level of significance. Hence it can be confidently stated that the two tests are equivalent.

4.3 Discriminative Index and Difficulty Levels.

The discrimination Inded and difficultu levels as a shown in Appendix F were calculated to give the final item analysis of the test. From Appendix F. one could see that about ninety (90) percent of the items in test I were within the standard required since they all have difficulty index ranging from 0.50 to 0.62. Only items number (1) and item number (8) could be said to be too easy. Since their F-value indicated that more examinees answered them correctly. Looking at test II only item (1) and (2) could be considered as easy questions while all the remaining ones were within the required levels of difficulty.

As for the discrimination power, Appendix F shows that all the items in test I with the exception of items (2) and (3) have very good disciminative powers because they are all from 0.40 and above. Turning to test II, the analysis looks the same because the items have what experts regard as good discriminative power except item number twenty (2) which Pophem (1981) described as marginal item usually needing and being subject to imporvement because it falls with 0.20 to 0.28.

It is interesting to note that the final item analysis shows some degree of improvement over that of the first try-out because most of the items in all the tests fall within 0.50 to 0.62 which is the good range of difficulty index. (See Appendix F.).

4.4 Comparative Analysis Of The Results By Location.

For this analysis the data collected were the pupils raw scores in the tests from both urban and rural schools. The analysis was done separately for pupils from both urban and rural in both form of the tests as show in table 4.3.

4.4.1 Comparison Of The Results By Locations

For this analysis the means and standard deviations of urban and rural schools were computed, the number of pupils from both urban and rural and the difference between the means

of both urban and rural were used to calculate the t-value. The investigations here were centred on whether pupils from urban who participated in the examination perform better than pupils from the rural schools.

Table 4.3: Means And Standard Deviations of Scores Of Urban and Rural Schools

(N = 772)

Test Set	Location	No. of Cases	Means	Stand. Dev.	T-Value	DF	2-tail prob.
Test I	Urban	179	20.37	5.03	15.23	393	.00
	Rural	217	12.87	5.33			
Test II	Urban	174	18.67	4.43	9.81	373	.00
	Rural	202	13.22	6.06			

4.4.2 Performance Between Urban Schools And Rural Schools

The null hypothesis tested here was:

H_o: There is no significant difference in the level of performance between Urban schools who took part in the examination and their counterparts in the rural areas.

H_i: The mean performance of Urban Schools is significantly different from that of the rural schools.

Where H_o: = null hypothesis

H_i: = alternative hypothesis.

The calculated t-value in test I was 15.23. This is greater than 1.960 the critical value of t at 0.05 level of significance, whereas the calculated t-value in test II was 9.81 which is greater than 1.960, the critical value of t at 0.05 level of significance. Taking the two results together, the null hypothesis was therefore rejected. On the basis of this, the alternative hypothesis was accepted. The conclusion is that the children from urban schools who participated in the tests performed better than their counterparts in the rural schools.

The finding might be connected with the fact that children in the urban schools have better learning facilities. These include among other things, better qualified teachers, good learning environment, greater concern on the part of the parents, and library facilities.

Table 4.4: Means and Standard Deviations of Boys and Girls Who Participated in the Examination.

Test Set	Sex	No. of Cases	Mean	Standard Deviations
Test I	Boys	259	16.32	6.48
	Girls	138	16.76	6.66
Test II	Boys	245	16.20	5.99
	Girls	130	14.74	5.94

The means and standard deviations of the results by sex as shown in table 4.4 indicates that boys and girls who participated in taking the examination have almost the same means and standard deviations in Mathematics Achievement Test. This actually dispels the assumption that boys perform better than girls in Mathematics. However, to further confirm this, a comparative analysis was taken to explain this point clearly. Thus to say a t-statistics was used to test the hypothesis testing the sex differences.

The study was also interested in finding out if there are any rural-urban differences in the students scores in the two tests with possible sex differences controlled. Hence the following sub-null hypotheses were then tested by means of t-test.

4.5 Comparative Analysis Of The Performance of Urban and Rural Boys. Urban and Rural Girls.

For this analysis, the means and standard deviations of boys and girls in each test, the number of both boys and girls, and the difference between the means of both sex were used to obtain the t-value. The investigations here we centred on whether boys from urban schools who took the test would perform better than boys from the rural schools, whether urban boys would perform better than urban girls who took the test and whether rural boys would perform better than rural girls.

Table 4.5: Means and Standard Deviations of Sex By Location

(N = 772)

Test Set	Location	Sex	No. of Cases	Mean	Stand. Dev.	T-Value	DF	2Tail
Test I	Urban	Boys	109	21.00	4.94	12.58	256	.00
	Rural	Boys	149	13.89	5.23			
Test I	Urban	Girls	70	20.65	5.17	8.52	135	.00
	Rural	Girls	67	12.82	5.58			
Test II	Urban	Boys	117	19.31	4.31	8.54	243	.00
	Rural	Boys	129	13.55	5.99			
Test II	Urban	Girls	58	17.39	4.42	4.96	128	.00
	Rural	Girls	72	12.61	6.18			

4.5.1 Comparison Between Boys From Urban School and Boys From Rural Schools In Test I.

The sub-null hypothesis tested here was:

H_o: There is no significant difference between the means performance of boys from urban schools and boys from rural schools who took part in the test.

H_i: The mean performance of boys from urban schools would be significantly different from that of boys from rural schools who took part in the test.

From table 4.5 it is evident that the calculated t-value in test I for boys is 12.58 which is greater than 1.960, the critical value of t at 0.05 level of significance. From this therefore the sub-null hypothesis was rejected. For this reason, the sub-null hypothesis was accepted. The conclusion here is that boys from urban schools who took test I perform significantly different from their counterparts from the rural schools.

4.5.2 Comparion Between Girls From Urban Schools and Girls From Rural Schools in Test I.

The sub-null hypothesis tested here was:

H_o: There is no significant difference between the means performance of girls from urban schools and their counterparts from the rural schools.

H_i: Girls from urban schools who took test I would perform significantly different from girls in the rural schools who took the same test.

From Table 4.5 above it can be seen that the calculated t-value is 8.52 which is greater than 1.960, the critical value of t at 0.05 level of significance. From this finding, the sub-null hypothesis was therefore accepted. The conclusion here was that girls from urban schools who took test I perform better than their counterparts from the rural schools.

4.5.3 Comparision Between Boys From Urban Schools and Boys from Rural Schools In Test II.

H_o: The sub-null hypothesis tested there was: There is no significant difference between the means performance of boys from urban and rural boys.

H_i: Boys from urban schools in test II would be significantly different from their counterparts in the rural schools.

Looling at table 4.5 once again one would see that the calculated t-value was 8.54 which is greater than 1.960, the critical value of t at 0.05 level of significance. The sub-null hypothesis was therefore rejected on the basis of this analysis. The conclusion here is that boys from urban

schools perform significantly different from their counterparts in the rural schools who took test II. It is interesting to note that the results of the analysis in test II is the same in test I.

4.5.4 Comparison Between Girls From Urban Schools and Girls from Rural Schools In Test II.

H_o: The Sub-null hypothesis tested here was:- There is no significant difference between the mean performance of girls who took test II from urban schools and those girls from rural schools.

H_i: The performance of girls from urban schools who took test II would be significantly different from their counterparts from rural schools.

Where H_o = sub-null hypothesis

H_i = sub-alternative hypothesis.

Going back to table 4.5, it could be observed that the calculated t-value was 4.96 which is greater than 1.960, the critical value of t at 0.05 level of significance. The null hypothesis was therefore rejected. The conclusion here is that girls from urban shool perform significantly higher than girls from rural schools who took test II.

The findings from these hypotheses might not be unconnected with the fact that boys from urban school concentrate more on their studies than boys in the rural school. In the urban schools there are lots of facilities which go along way to facilitate learning in the townships. It was also observed that some boys in the urban schools have some part-time teachers who coach them at home in addition to what they schools offer to them. Again most of the parent in the urban schools are educated ones and majority of them do help their owh children. Turning to the rural schools one would see the boys are at advantage because they do not have all those facilities mentioned earlier, Another reason that makes boys from urban schools better than their counterparts in rural schools is that boys in urban schools have enough time for their studies while their friends in the rural areas do heavily engage themselves in farming. This is very true because some of the Headmasters in the rural schools complained to the researcher that this time of the year children don't come to schools regularly. This inconsistency in coming to schools by the rural boys is definately the cause of their backwardness.

Turning to the hypotheses that tested the relationship between girls from the urban schools and their counterparts from the rural school, it can easily be seen that the

performance of girls from urban is significantly higher than girls from rural schools. The reasons here are, first the girls from urban school enjoyed all those facilities mentioned whereas their counterparts in the rural schools are at a disadvantage. Another reason connected with this development is that the researcher discovered that girls from rural areas to engage themselves heavily in petty trading and they come school late or sometimes they don't even come. This is not the case with their friends from urban schools who don't usually engage in road-side trading.

From the findings of these four sub-null hypotheses it is evidently clear that there is the need for separate norms for boys and girls which take into consideration their environmental backgrounds. These findings again support the findings of Ikeotuonye (1982) and Jahun (1985) who both advocated for separate norms for boys and girls so at to cater for their environmental backgrounds.

Table 4.6: T-Test of Sex Differences on The VAMAT Urban Boys And Girls.

(N = 772)

Test Set	Location	Sex	No. of Cases	Mean	Stand. Dev.	T-Value	DF
Test I	Urban	Boys	109	21.00	4.94	0.449	177
	Urban	Girls	70	20.65	5.17		
Test I	Rural	Boys	149	13.89	5.23	1.350	214
	Rural	Girls	67	12.31	5.58		
Test II	Urban	Boys	117	19.31	4.31	1.740	173
	Urban	Girls	58	17.39	4.42		
Test II	Rural	Boys	129	13.55	5.99	1.045	199
	Rural	Girls	72	12.61	6.18		

The study is also interested in finding out if sex of students has any effect on their VAMAT scores. The t-test technique was therefore computed to find out if there is any such effects.

From the table above it is found that in the case of urban boys and girls their scores of test I are similar. This means there is no significant difference between the mean performance of boys urban boys and urban girls in test I

Also it could be seen from the table that the calculated t-value of 1.350 for rural boys and girls, is less than the critical value of t at 0.05 level of significance.

We can therefore accept the null hypothesis. This means that there was no a significant different between the mean performance of the two groups.

Turning to test II, the table shows that for Urban boys and girls, the sub-null hypothesis tested here is accepted at 0.05 level of significance. Again for the rural boys and girls the above hypothesis is accepted at 0.05 level of significance which shows that mean performance of rural boys and girls is similar.

From the foregoing it is evident that the performance of boys and girls in all the two test is similar.

4.6 Reliability Estimate

The pupils raw scores in each test were corrected as indicated in chapter 3 using Person Product Moment Correlation formula:-

The table below gave the reliability coefficient of the two tests by the sex and location.

Table 4.7: Equivalent Reliability Coefficients of The Two tests by Sex and Location.

(N = 166)

	Urban Boys	Rural Boys	Urban Girls	Rural Girls
	.79	70	.75	.73
N	50	70	20	26

Looking at the table 4.7, it would be of interest to not that the four correlation coefficients are within .70's. A high correlation indicated that both forms of test were providing similar results and therefore werecertainly reliable samples of the general areas of content being measured. On the basis of this, the above analysis woundn't be said to be high or to low, it was certainly within the average acceptance. However, since these test-sets were not meant for standardization the obtained reliability coefficients could be said to be good enough.

We could observethat the reliability coefficients of the urban children was higher than that of the rural children. This could be attributed to the fact that the urban children had a more consistent scores than their counterparts in the rural schools. Another reason might be that the test wiseness of rural children was less than that of the urban children. However, from these results we could infer that the two tests were paralled and interchangable.

4.7 Standard Error Of Measurement

The table below shows the standard error of measurement SEM for sex and location. As you all know we take the standared error into account when interpreting individual test scores. For this reason the standard errors of measurement for boys, girls and location was computed.

Table 4.8: Standard Error Of Measurement

Test Set	Total	Boys	Girls	Urban	Rural
SEM	.33	.40	.56	.38	.36
Test I					
N	396	159	137	179	217
SEM	.31	.38	.52	.34	.43
Test II					
N	376	245	131	174	202

Looking at the table it can be observed that the confidence band for the boys who took test I is ± .40 while that of test II is ± .38. The analysis also revealed that the confidence-band for the girls in test I and test II were ± .56 and ± .52 respectively. The confidence band for the urban children in both test I and test II were ± .38 and ± .34 respectively, whild that of the rural children were obtained as ± .36 and ± .43 respectively.

4.8 Concurrent Validity

The pupils raw scores in the test were correlated with their observed scores using Pearson Product Moment Correlation formula. The results of the correlation showed that test I has a validity coefficient of .88 while test II has a validity coefficient of .86.

Table 4.9: Concurrent Validity Coefficient of The Test Sets.

Test Set	Validity Coefficient
Test I	0.88
Test II	0.86

From the above data, one can see that the validity coefficient of the Test I is some-what perfectly positive. While the one in Test II is also positive and was abit less than the one obtained in Test I. Obviously, the nearer a validity coefficient approaches 1.00, the happier we are with it because larger validity coefficient indicates greater accuracy in predicting from one variable to another. With this therefore we could say that the finding showed that the scores obtained in the two sets of tests were really true scores of the children.

Looking at the validity coefficients of the two tests it could be observed that the two coefficients were significant at 0.05 level of significant, and this really confirmed that the tests were valid. This certainly answered the question raised in chapter one.

Chapter 5

SUMMARY, FINDING AND CONCLUSION, AND RECOMMENDATION.

5.1 Introduction

In this chapter, the emphasis would be on the summary of the purpose of the research, literature review, methodology as well as the analysis of the data. And since the Data of this research has been documented and analysed in chapter 4, this chapter will also discuss on the findings and make conclusions. The manual information and recommendations were also looked into.

5.2 Summary

This section discusses a summary of the research purposes, related literature, methodology, hypothesis and finding.

5.2.1 Purpose Of The Study

Since it has been outlined in chapter one, that the study was motivated by the desire of the research to find out some major problems facing primary schools Mathematics Achievement Test. The purpose of the study was to construct and validate primary three Mathematics to revise and edit the test items and finally come out with validated primary three Mathematics test battery called Validated Mathematics Achievement Test (VAMAT).

The study also tried to answer some questions which were outlined in chapter one sub-section 1.4 and based on these questions some hypotheses were formulated. These hypotheses were throughly tested in chapter 4. The findings of the tested hypotheses were also recorded.

5.2.2 Related Literature

The review of literature which focused on what were discussed in chapter 2 was hereby summarized. Since the research concerns the evaluation of pupils at the lower level of primary schools, the researcher reviewed at length what psychologists and mathematics educators said about children at this tender a age of their development. The review therefore looked at what those experts advocated at the concrete operational stage since the age of the population under study coincide with the children's cognitive development.

The content of the primary three Mathematics were looked into with the view to highlighting what mathematics educators recommended to be done my Mathematic teachers.

Since evalustion is the major focus of this study the researcher reviewed evaluation in general thereby focussing more on the concept of evaluation model which looked at cognitive domain of the educational objectives. The review went further to talk on evaluation procedures that are relevant to the study. The technique of test construction were thoroughly examined in the review.

The validation aspect was looked into. Here, the review highlighted the modality of getting difficulty index and discriminative power. The psychometric qualities were reviewed. In this section the validity as well as the reliability of the tests were looked at.

5.2.3 Data Collection Procedures

The population under study are the primary three pupils in Zaria Local Government Area. Out of the fifty (50) primary schools within Zaria Metropolitan, five (5) of them were stratified sample for the execution of the first try-out. For the final try-out, the entire Local Government was categorised into:- (a) Districts and (b) Public and Private schools in order to ensure that each characteristic was fully represented in the sample. The population was therefore stratified as follows:-

a) The Urban Public Schools
b) The Private Public Schools
c) Rural Public Schools
d) The Rural Private Schools.

Having stratified the population into strata, seven schools were taken from the urban schools, two from urban private schools, eleven from rural public schools and one from rural private schools.

The instrument used for the collection of data was test materials. It was therefore because of this, the methodology include a detailed process of test construction.

5.2.4 The Hypotheses And Instrumentations

Particular statistical instruments were used to test the hypothesis and this was done in the following manner. The analysis include the finding of means, standard deviations, difference of two means by the t-statistic, 0.05 level of significance and correlation formula of Pearson Product Moment.

To test the first null hypothesis (H_o:1), for example a t-statistic was used to find out whether or not there was any significant difference between the two parallel test papers.

5.3 Findings And Conclusion

The statistical analysis of the data in this research was discussed in chapter 4. The following were the major findings and conclusions. Looking at the means and standard deviations of the test sets, it was discovered that their values in each case were found to be relatively the same.

5.3.1 Hypothesis 1

The first null hypothesis was tested using t-statistics to determine whether there is a significant difference in the

variability of scores between the scores of the pupils who took both forms of the tests. The null hypothesis was accepted at 0.05 level of significance. The findings was that there is no significant difference in the variability of scores of the students who took both forms of the tests.

5.3.2 Hypothesis 2

The null hypothesis was rejected. This shows that a significant difference was found between the urban schools and rural schools. The finding here shows that children from urban schools who participated in the test peform better than those children who came from the rural schools. One can conclude that this development came as a result of the fact children in the urban school have better learning facilities than their rural counterparts.

5.3.3 Hypothesis 3

The main null hypothesis was divided into two sub-hypothesis:- (a) There is no significant difference between the means performance of boys and girls in test I. (b) There is no significant difference between the means performance of boys and girls in test II.

The first sub-hypotheses was rejected at 0.05 level of significance. The finding indicates that there was a

significant difference between the performance of urban boys and that rural boys in test I. The second sub-hypothesis was rejected at 0.05 level of significance. The conclusion here was that urban girls who took test I perform better than rural girls who took the same. The analysis of the last two sub-hypothesis indicated that boys from urban schools perform better than their counterparts in the rural areas. The same results extended to show that girls from urban schools perform significantly higher than girls from the rural areas.

5.4 Reliability Coefficients

Turning to table 4.7 it is of interest to note that the four correlation coefficients were within the .70's. This indicated that there is a high positive correlation between the groups, and the remaining two have moderate positive correlation. It can also be observed that the reliability coefficients of the two sets of tests were very close, this indicated that the tests were reliable. The conclusion here is that the two sets of tests can be said to be parallel and interchangable.

5.5 Standard Errors Of Measurement

It can be seen from the table 4.8 in chapter 4 that the margin of errors surrounding the individual's test score is very low for all the variables.

Gronlund (1981) said that if the standard error is small the band of scores is small and we have greater confidence that our obtained score is a dependable measure of the characteristics.

The findings indicate that the confidence band for the boys who took test I is ± .40 while that of test II is ± .38. The analysis also revealed that the confidence band for the girls in test I and test II were +- .56 and +- 52 respectively.

The confidence-band on the urban children and rural children were ± .38 and ± .34 respectively.

The conclusion here is that the examinees scores for boys in test I and test II lie plus or minus .40 and .38 respectively. Thus their true score in test I is ± .40 while in test II is ± 38. The examinees scores for girls in test I and test II lie plus or minus .56 and .52 respectively i.e. their true scores in test I was ± .56 while in test II was ± 0.52. The examinees scores for the urban children who took test I and those who took test II lie plus or minus 0.38 and ± 0.34 respectively, that is to say that the true scores of the children from urban school were ± 0.36 and ± 0.33 respectively. The finding also

showed that the examinees scores of the rural children in test I and test II were ± .36 and ± .39 respectively. This indicated that the examinees true scores in both test I and test II were ± .36 and ± .39 respectively. On the basis of these findings one was therefore safe to claim that the tests did infact measure the true performance of the individuals and this further strenghten the reliability of the tests for all the variables.

5.6 Validation Process

From available literature, it was found that for one to validate test materials, the following had to be thoroughly treated:-

1. Psychometric qualities
2. Content validity
3. Concurrent validity
4. Reliability estimates.

The item analysis conducted in chapter 3 as could be seen in Appendix 'C' served as a frame work for selecting the items that were used for the final try-out. This was because it yielded difficulty index and discrimination power that were within the acceptance limit. The result of the final try-out was again subjected to another item

analysis. For example, in Appendix 'F', one would be convinced that about (90) percent of the items in both test I and test II were within 0.50 to 0.62. Since the difficulty Index and discrimination power of the two sets of tests were within an acceptable range it could be seen that this part of validation process was achieved.

In order to ensure content coverage for the test, the whole curriculum of primary 3 as shown in Appendix 'A' was gone through by making sure that both the topics and behaviour levels were looked into when constructing the table of specification for the test as indicated in paragraph 3.5.

To ensure reliability of the tests an equivalent form reliability coefficient was calculated and the result obtained was found to be significant at 0.05 level of significance, as could be seen in paragraph 4.8.

Concurrent validity was looked into in order to ensure that the observed scores were really the true scores of the pupils. The validity coefficient obtained as shown in paragraph 4.9 were significant at 0.05 level of significance.

From the foregoing discussion, it could be observed that the test battery prepared by the researcher for the use of primary 3 pupils could be said to be a validated test materials.

5.7 The Use To Be Made Of (VAMAT)

The tests have been designed to serve the following functions.

1. Diagnostic Function:

As it has been outlined in chapter One under section 1.3, the tests were meant to be used in primary 3 class so as to diagnose some of the problems pupils are facing in Mathematics. This can be done easily by checking the script a child answers and spot out the items he fails. With this process therefore some areas where the child has a problem can be identified and quick remedial work can follow. It is hoped that this test material can go to our primary schools so that teachers can use them to diagnose their pupils problems.

2. Continuous Assessment:

We all know that with the introduction of the new Policy on Education, the system demand that Continuous Assessment should be given a high recognition. This system is not only restricted to secondary schools alone, primary schools are also expected to make best use of the system. A standard is therefore expected in all the schools. This is very difficult to exist if a uniform

material was not developed. It is therefore along this line that this test battery could be of use. It is hoped that VAMAT can be used to check students performance across the schools at the end of the primary 3 level of education.

5.8 Recommendation For Further Research

The data obtained, findings and conclusions so made from this study calls for a serious thought. It therefore because of this, the following recommendations are being suggested.

First the pupils performances differ seriously. The performance of the pupils from the rural primary schools was very poor compare to that of their counterparts from the urban schools. A close look should be given to the rural schools which the study found that they lack so many educational facilities. Local Government Authorities should see that all the educational facilities are shared equally so that the rural schools can get its own share.

Secondly, the students poor performance in general shows that there is a serious problem which needs to be investigated further. Along this line therefore a diagnostic work should be carried out in order to spot out the areas of difficulty.

From the items analysis it was discovered that addition with carrying, multiplication with carying and division were the major areas the students had problems. A quick remedial work can therefore be done to remedy the problems.

Thirdly, the use of Hausa as a medium of instruction should be strictly adhered to. This is because it was discovered that the major problems the pupils have at the lower level of primary school is the language. From the study it was also found that the pupils found it difficult to read the examination written in Hausa. Teachers of Hausa language should work hard to improve the pupils knowledge in this direction if meaningful learning can take place.

Fourthly, the test result can be subjected to further statistical analysis in order to check the difficulty level and discriminating power. Those of them proving very difficulty or showing power discriminating can be reviewed or replaced.

Fifth, this research may be expected to cover the whole Kaduna State or even the whole country.

Sixth, a similar study may be conducted to cover the other classes or primary schools. This is because it is believed that an extension of this research to cover the

standardization of the primary six Mathematics test materials will be of great help to examination bodies such as National Teachers Institute and West African Examination Council.

On the other hand teachers in primary schools should make use of this research materials. This will provide them with first hand information as to how they should construct and validate their mathematics test materials.

Lastly, teachers in primary schools are recommended to use the test battery produced at the end of this study in order to maintain a uniform continuous assessment document to cut across all the primary schools in the Local Government. Also Educational Resource Centres are expected to make use of this research material to conduct refresher course for the primary schools teachers on how to construct and validate Mathematics test materials.

BIBLIOGRAPHY

Adams, G.S. & Torgerson, T.L. (1966). Measurement and Evaluation in Educational Psychology and Guidance, Holt Rinehert and Winston, Pg. 110-125.

Alan, S. (1971). The effect of answer sheet format on Performance by culturally disadvantage fourth grade Elementary School Pupils, Journal of Educational Measurement, 8(4): 191-195.

Anastasi, A. (ed) (1976). Psychological Testing, New York, the Macmillan Company, Pg. 105-125.

Avital, S.M. and Shettleworth, S.J. (1960). Objective for Mathematical Learning some Ideas for the Teachers. The Ontario Institute for Studies in Education.

Balow, I.H. (1964). Reading and Computation Ability as Determinants of Problem Solving. The Mathematics Teacher, 11(4): 160-164.

Best, J.W. (1970). Research in Education. 2nd Ed. New Jersey; Prentice-Hall Inc., Englewood Cliffs, Pg. 25-40.

Bloom, S.B. (Ed) (1956). Texanomy of Educational Objective: The cl$_3$ssification of Goals, Handbook I: Cognitive Domain, New York, Mckay, Pg. 240-265.

Bloom, et al (1971). Handbook Handbook on Formative and Summative Evaluation of Students Learning, New York: McGrow-Hill. Pg. 195-205.

Bormuth, J.R. (1970). On the Theory of Achievement Test Items. Chicago, University of Chicago Press, Pg. 45-65.

Brown, F.G. (1979). Principles of Educational and Psychological Testing. Holt, Rinehart and Winston, Pg. 100-120.

Brown, F.G. (1983). Principles of Educational and Psychological Testing. 3rd Edition, C.B.S. College Publicating, New York, Pg. 149-165.

Bruner, J.S. (1963). The process of Education, Vintage Books a demension of Random House New York. Pg. 129-151.

Capps, L.R. (1963). A Comparison of the Common Donominator and Inversion, Method in Division of Fraction. Journal of Educational Researh, 12(3: Pg. 156-168.

Clark, C.A. (1968). The use of Separate Answer Sheet in testing show learning pupils. Journal of Educational Measurement, 5(2) Pg. 61-64.

Costello, J. (1970). The effect of pretesting and examiner Characteristics on test performance of young Disadvantaged Children, Preceeding of the 78th Annuals Convention of American Psychological Association, 8(2): 309-310.

Davis, F.B. (1975). Lowa Students Reading Tests. Journal of Educational Measurement 3(1): 57-60.

Ebel, R.L. (1979). Essential of Educational Measurement, Prentice-Hall Inc., Eglewood Cliffs, New Jessey 3rd Edition, Pg. 100-168.

Fakuode, R.A. (1981). Teaching Arithmetic and Mathematics in the Primary School, Ibadan University Press.

Fergeson, G.A. (1981). Statistical Analysis in Psychology and Education, McGraw Hill Inc., Fifth Edition, Pg. 177-178.

Fox, D.J. (1969). The Research Process in Education; New York, Holt Rinehart and Winston, Inc.

Fred, T.W. (1967). Evaluation as Feedback and Guide. Association of Supervision and Curriculum Development, Washington D.C., 127-130.

Furst, I.J. (1976). Constructing Evaluation Instruments, New York, Pg. 125-140.

Gerberich, J.R. (1962). Specemen Objective Test Items; A Guide to Achievement Test Construction, New York, Longman Greson, Pg. 155-160.

Gerberich, J.R. et al (1962). Measurement and Evaluation in Modern Schools. New York Mekay, Pg. 148-159.

George, T. (1971). Validity of the SAT for Blacks and Whites in Thirteen Integrated Institution. Journal of Educational Measurement, 8(4). Pg. 245-251.

Gibb, E.G. (1956). Children's Thinking in the Process of Subtraction, Journal of Experimental Education 25(2): Pg. 71-80.

Glass, G.U. et al (1970). Statistical Methods in Education and Psychology, Prentice-Hill Inc.

Goslin, D.A. (1967). Teachers and Testing. Rusell Sage Foundation.

Gray, R.F. (1965). An Experiment in the Teaching of Introductory Multiplication. The Arithmetic Teacher, 12(6): Pg. 121-123.

Gronlund, N.E. (1981). Measurement and Evaluation in Teaching, MacMillan Publishing Co., Inc. New York, Pg. 65-90.

Grossnickle, et al (1983). Descovering Meanings in Elementary School Mathematics, Holt, Rinnechart and Winston. Mathematics Learning in early Childhood. National Council of Teachers of Mathematics, Pg. 240-265.

Hammill, D.D. (1971). Evaluating Children for Instructional Purposes Academic Theraphy.

Hall, K.D. (1967). An experimental Study of two Methods of Instruction for Mastering Multiplication Facts at the Third grade level. Unpublished Doctorial Dessertation, Duke University, Second Handbook of Research on Teaching.

Harvey, M.A. (1966). Children's responses of two types of Multiplication Problems. The Arithmetic Factor, 13(8): Pg. 288-292.

Hills, J.R. (1981). Measurement and Evaluation in the Classroom. Columbia, Ohio Merrill, Pg. 128-142.

Ikeotuonye, A.I. (1982)). A Validation study of the Defferential Aptitude Tests in Kaduna State of Nigeria: Implication for Guidance, Unpublished Ph.D. Theses, University of Wales.

Jahun, U.I. (1985). An Attempt at Standardization of Junior Secondary Three Algebra Test Using the Contemporary Mathematics Curriculum in Kano State of Nigeria. Unpublished M.Ed. Theses.

Kalejaiye, A.O. (1982). Evaluation of Achievement in Secondary School Mathematics. Abacus, The Journal of the Mathematical Association of Nigeria, Volume 16.

Kurkani, S.S. (1968). Constructing Equivalent Achievement Survey Tests to be used for Different Student Populations in India. Development in Educational Testing, Volume I.

Layman, H.B. (1963). Test Scores and What they Mean Eglewood Cliffs N.J. Prentice-Hill, Pg. 111-121.

Lundaine, A.A. (1963). Instrument of Media of Instruction In. N.L. (ed) Handbook of Research on Teaching. Rand Mcnally & Company Chicago U.S.A.

Mechatchy, J.H. (1953). Another Measure of the Difficulty of Addition Combinations. Educational Research Bulletin, 10(3): Pg. 100-120.

Martuza, V.R. (1977). Applying Norm-Referenced and Critirion Referenced Measurement in Education. Eglewood Cliffs, N.J. Prentice-Hall.

Marshall, J.C. & Hules, L.W. (1971). Classroom Test Construction, Addison-Wesley Publishing Company Inc., Pg. 146-158.

Mehrens, W.A. & Lehmann, L.I. (1975). Measurement and Evaluation in Education and Psychology. New York, Holt Rinehart Winston Inc. Pg. 313-345.

Merwin, J.C. (1965). "Constructing Achievement Test and Interpreting Scores" In Evaluation in Mathematics: National Council of Teachers of Mathematics.

Millmer, J. (1980). Reporting Students' Progress, A Case for Criterion Referenced Marking System Phi Della Kappan, Volume 11, No. 4 11(4): 126-140.

Maller, et al (1965). An Analysis of test Wiseness. Journal of Educational and Psychological Measurement. 25 (11): 157-169.

National Policy on Education (1981). Lagos, Federal Ministry of Education.

Obioma, G.O. (1985). The Development of Diagnostic Mathematics Test for Junior Secondary School. University of Nigeria, Nsukka.

Ohuche, R.O. & Akyu, S.A. (1977). Testing and Evaluation in Education. African Educational Resources. Sky-way Press Limited, Ikeja.

Ohuche, R.O. (1982). Assessment of Cognitive Outcomes. A Paper presented at the National Primary Science/ Mathematics Workshop held at Zaria.

Olomolaiye, F. (1987). A questionnaire of Children's Attitude to school, WAEC Seminar papers, 1987.

Onwana, O.C. (1981). Educational Measurement for Teachers, Nelson Africa.

Osborne, A.R. (1960). The Effect of Two Instructional Approaches in the Understanding of Subtract by Grade Two Pupils. Unpublished Doctorial Dessertation, University of Michigan.

Osualu, E.C. (1982). Introduction to Research Methodology. Africana - FEP Publishers Limited.

Piaget, J. and Inhelde, B. (1959). The growth of Logical Thinking, London; Routledge & Kegan Paul, Pg. 249-264.

Popham, W.J. (1981). Modern Educational Measurement. Prentice-Hall, Inc., Eglewood Diffs, Jessey, Pg. 292-340.

Popham, W.J. (1978). Criterion-Referenced Measurement. Eglewood Diffs, J.N. Prentice-Hill, Pg. 310-340.

Rummers, H.H. et al (1960). A Practical Introduction to Measurement and Evaluation, New Dehi, University Book Stull.

Stanley, J.G. & Hopkins, K.D. (1972). Educational and Psychological Measurement and Evaluation. Eglewood Diffs Prince-Hill Inc.

Sex, G. (1968). Empirical Foundation of Educational Research. New Jessey Prentice-Hall, Inc. Eglewood Diffs.

Scott, L. (1962). Children's Concepts of Scale and the Subtraction of Fractions. Journal of Research on Teaching 9(4): Pg. 20-35.

Second Handbook of Research on Teaching American Educational Research Association, 1973.

Swenson, E.J. (1944). Difficulty Rating of Addition facts as Related to Learning Method. Journal of Educational Research. 13(5): 169-174.

Thorndike, R.L. & Hagan, E. (1977). Measurement in Psychology and Education. New York: Wiley, Pg. 149-170.

Tuckman, B.W. (1978). Measuring Educational Out-comes. Harcourt. Brace Jooravovick Inc.

Wallace, G. & Larsen, S.C. (1978). Educational Assessment of Learning Problems. Testing for Teaching Allyn and Bacon, Inc.

Wason, P. (1961). Response to affirmative and negative binary statements. British Journal of Psychology. Pg. 133-142.

Wilson, J.W. (1971). Evaluation of Learning in Secondary School Mathematics. McGraw-Hill Book Company.

Wood, R. (1969). Multiple-Choice: A State of the Art, Sglongh. National Foundation of Educational Research. 10(4): Pg. 169-176.

Zern, D. (1967). Effects of Variation in Questions Praising on true-false answers by grade-school children. Psychological Report, 20(3): Pg. 149-161.

APPENDIX A

OUTLINE OF "PRIMARY EDUCATION MATHEMATICS CURRICULUM" (FED. MINISTRY OF EDU. 1979)

CLASS	NUMBER AND NUMERATION	BASIC OPERATIONS	MEASUREMENTS	PRACTICAL & DESCRIPTIVE GEOMETRY	EVERYDAY SATISTICS
1	count to 100 read, write to 40 tens, units(TU) half, quarter	add: sums to 20 subt: to 20-10 add 2-digits(TU) to 40, without "carrying" easy "missing number probs	coins, easy arith in kobo compare, order lengths, wts (no formal units) general times of day	-	-
2	count to 200 read, write to 100 odd, even; 2s, 5s, 10s more half, quarter three-quarters	add, subt. TU add three 2-digits renaming ("carry" and "borrow") mult. as repeated addition	coins, notes: use, make change to ₦ 1 metre, cm compare areas, wts, capacities (without units) time to hr, half-hr days of week	identify cuboids, cylinders, spheres rectangle, circle, square	-
3	count, write to 1000 HTU fractions of numbers (to 60) and shapes simple comparing of fractions	add, subt HTU add, subt fractions with same denom mult to 9 x 9 mult 2-digit by 1-digit multiples div up to 48 by 2,3,4,5,6 with no remainder factors to 48	change to ₦ 10 shopping to ₦20 mult. money (easy) estimate lengths meas. perimeters idea of area meas. litre, kg minutes, dates (month, date)	faces, corners, edges of cubes, cuboids "square corner" (much practical familiarization with geometry properties)	-
4	count in thousands apply counting in groups (5s, 12s) ThHTU Roman num. to L.C proper, improper fractions dec.frac to .01 equiv.fractions	add, subt ThHTU & all frac.ded mult by TU to 50 perfect squares, sq.rts: 1, 4, 9, ..., 100 div by units, div by 10, 20, 30, 40, 50 LCM, HCF (simple) estimates "missing Nos."	calc. money profit, loss arithmetic with length, wt, capac. am, pm in time area of rectangle (lengthxbreadth)	folding symmetry special properties of square, ract. types of triangles drawing shapes	pictogram: read, draw mode (most common value)
5	millions percents from decimal ideas primes less than 100 intro to ratio	mult by HTU of ₦ as mult. dec. xwholes perfect squares, sq.rts to 400 divide by TU (easy) more LCM, HCF calc. ratio, % harder "missing number probs	Naira, other money w.r.f., $, £ profit, loss simple interest commission, disct. perim., circumf. rt. triangle area ideas, formula of volume litre-1000cm^3 word probs-wt dist., time, speed degrees Celsius	symmetrical shape right, acute, obtuse angles horizontal, vertical north, south, east, west	Bar graphs read, make
6	revision word problems ratio, proportion (direct, inverse)	revise add, subt, mult, div div dec. by whole indices ($use in factorisation) word probs; ratio% order of operation use of letters in open sentences (pre-algebra)	taxes, shares ready reckoner use of Pythagoras perim-area area of irregular shapes by rect. & triangles volume: prisms probs: Litre, wts. minute, second time-tables	parallel perpendicular quadril., trapez., parallelogram, rhombus	collect, present data tally, tables graph, charts mean (average)

APPENDIX B

TEST I

PRIMARY 3 MATHEMATICS TRIAL TEST

INSTRUCTIONS: For each question, three answers are shown. Find which which is the correct answer. Tick the correct answer in the answer sheet.

Time Allowed: 45 Minutes.

1. Write down the missing numbers:-

   5, 6, 7, ☐, ☐, ☐, ☐, 12 13

   (a) 4, 8, 9, 10

   (b) 8 9 10 11

   (c) 1, 2, 3, 7

2. Write in figures

   Thirty :-

   (a) 13

   (b) 14

   (c) 30.

3. Write in words

   90:-

   (a) Ninety

   (b) Nine

   (c) Nineteen.

4. Write in ordinary numbers

   XC:-

   (a) 80

   (b) 90

   (c) 110

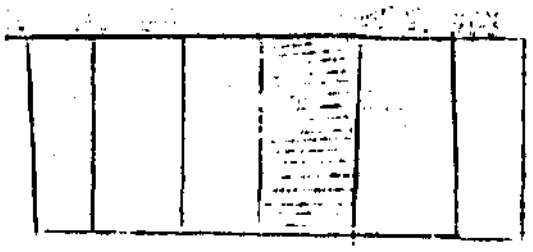
5. What number does the (3) stands for in the 7(3)8

   (a) 2

   (b) 30

   (c) 10.

6. What fraction of the rectangle is shaded



(a) 1/7

(b) 1/5

(c) 1/6

7. 1/6 = $\frac{\quad}{12}$

(a) 2

(b) 6

(c) 12

8. What is 3/8 of 16

(a) 6

(b) 2

(c) 5

9. Add the following ½ + ¼ =

(a) ¾

(b) ¾

(c) ½

10. 349
+ 406

(a) 745

(b) 755

(c) 756

11. 27, 32 and 57

(a) 106

(b) 116

(c) 28.

12. Write down the even numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

(a) 1 3 4

(b) 2, 4 6, 8, 10

(c) 9, 10, 6.

13. Write down the factors of 24

(a) 6 and 4

(b) 8 and 4

(c) 6 and 2

14. Subtract 47 − 29 =

(a) 18

(b) 76

(c) 15

15. 345 − 232 =

(a) 113

(b) 103

(c) 349.

16 63 − 49 =

(a) 14

(b) 19

(c) 10

17. Subtract 455 from 650

(a) 125

(b) 195

(c) 659

Write down the number in the box

118 + ☐ = 200

(a) 20

(b) 38

(c) 2.

18. 105 x 15 =

(a) 1545

(b) 525

(c) 705

19. Multiply 17 by 4 =

(a) 48

(b) 61

(c) 68

20. ▱ x 12 = 144

(a) 10

(b) 12

(c) 14

21. 10 x 3 x 4 =

(a) 30

(b) 22

(c) 120.

22. 4 ⟌ 24 =

(a) 8

(b) 12

(c) 6

23. Find the remainder 17 ♣ 3

(a) 3

(b) 1

(c) 2

24. Measure the length in centimetres

(a) 9cm

(b) 9.5cm

(c) 10cm.

25. What is the time?



(a) Half past five

(b) A quarter past five

(c) Three O'clock.

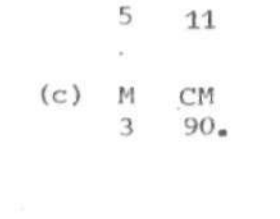
26.
$$\begin{array}{r} 2.64 \\ +\ 5.65 \\ \hline \end{array}$$

(a) ₦8.29

(b) ₦3.01

(c) ₦4.56

27. Find the area of



(a) $11cm^2$

(b) $30cm^2$

(c) $1cm^2$

28. How many faces has a cube?

(a) 8

(b) 6

(c) 9

29. Change to naira 209K =

(a) ₦2.90

(b) ₦2.09

(c) ₦209.00

30.
$$\begin{array}{rr} M & CM \\ 2 & 39 \\ +\ 1 & 72 \\ \hline \end{array}$$

(a)
$$\begin{array}{rr} M & CM \\ 4 & 11 \end{array}$$

(b)
$$\begin{array}{rr} M & CM \\ 5 & 11 \end{array}$$

(c)
$$\begin{array}{rr} M & CM \\ 3 & 90. \end{array}$$

PRIMARY 3 MATHEMATICS TRIAL TEST

INSTRUCTION:- For each question, three answers are shown. Find which is the correct answer. Tick the correct answer in the answer sheet.

Time Allowed: 45 Minutes.

1. Write in figures

   Twenty = -- (a) 10 (b) 20 (c) 30

2. Write in words

   12 = -- (a) Twelve (b) Twenty (c)Eleven

3. XXXIV = -- (a) 24 (b) 35 (c) 34

4. What number does the digit (9) Stands for in 74 (9)?

   (a) 2 (b) 9 (c) 90

5. What fraction of the circle is shaded



   (a) ¼ (b) 2/7 (c) 2/5

6. $\frac{2}{3} = \frac{\quad}{9}$ (a) 3 (b) 5 (c) 6.

7. What is 4/5 of 40 (a) 30 (b) 40 (c) 32

8. Add the following:- 1/5 + 2/5

   (a) 3/5 (b) 2/5 (c) 2/25

9. 58 + 69 = (a) 28 (b) 117 (c) 127

10. $\begin{array}{r} 654 \\ +\ 462 \\ \hline \end{array}$ = (a) 1161 (b) 1116 (c) 1016.

11. 340 + 47 + 8 = (a) 395 (b) 55 (c) 387.

12. Add forty nine and seventy eight.

    (a) 127 (b) 64 (c) 107.

13. Write down the number in the box. $\square$ + 109 = 118

    (a) 9 (b) 17 (c) 19

14. Simplify (16 + 24) + (20 + 19)

    (a) 79 (b) 49 (c) 39

15. 978 - 609 = (a) 369 (b) 269 (c) 279

16. Subtract 45 from 105 =

(a) 40 (b) 60 (c) 75

17. Simplify (400 - 350) + (2504 + 50) =

(a) 300 (b) 50 (c) 350

18. Find the difference between 81 and 18

(a) 99 (b) 73 (c) 63

19. $\frac{105}{\times 5}$ = (a) 525 (b) 110 (c) 705

20. Multiply 23 by 4 = (a) 48 (b) 61 (c) 92

21. Simplify (8 x 4) - (16 x 2) (a) 0 (b) 64 (c) 32

22. Add ₦ 5.47 and ₦3.44 = (a) ₦8.81 (b) ₦8.91 (c) ₦8.03

23. 3 ÷ 30 = (a) 10 (b) 3 (c) 27

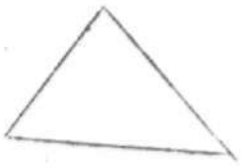
24. How many grammes are there in 2½kg?

(a) 200 (b) 250 (c) 2½.

25. If the cost of one hen is 49k. What is the cost of 2 hens.

(a) 48k (b) 98K (c) 96k

26. Write down the name of the shape below.



(a) Triangle

(b) Square

(c) Rectangle

27. What time is it?

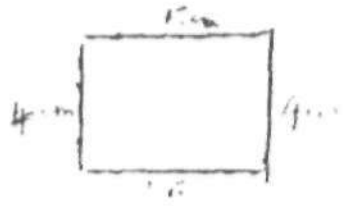


(a) 20 minutes past One

(b) Half past One

(c) One O'clok.

28. Find the perimeter of the figure below



(a) 14cm

(b) 20cm

(c) 18cm

29. Write in centimetres 1m 32cm

(a) 123cm

(b) 132cm

(c) 100cm

30. Change to Kobo ₦1.35

(a) 1351Kobo

(b) 135 Kobo

(c) 105 Kobo

TEST III

PRIMARY 3 MATHEMATICS TRIAL TEST

INSTRUCTIONS:- For each question, three answers are shown. Find which is the correct answer. Tick the correct answer in the answer sheet.

Time Allowed: 45Minutes

1. Write in figures Eighty - Eight =

   (a) 18 (b) 88 (c) 80.

2. Write in words

   22 = - (a) Twenty (b) Twelve (c) Twenty-Two

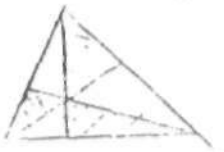
3. Write down the odd numbers

   1, 2, 3, 5, 6, 7, 8, 9, 10

   (a) 1, 2, 3, 4, 5,

   (b) 3, 5, 6, 9

   (c) 3, 5, 7, 9

4.  = (a) 5/6 (b) 3/8 (c) 3/6

5. $\frac{1}{5} = \frac{5}{\square}$ (a) 20 (b) 10 (c) 25.

6. What fraction is 6 of 24

   (a) 1/3 (b) 1/4 (c) 2/3

7. Add $\frac{1}{5} + \frac{3}{5}$

   (a) 4/5 (b) 4/6 (c) 3/25

8. Subtract $\frac{1}{2} - \frac{1}{4}$ (a) 1/4 (b) 1/2 (c) 2/3

9. Add the following

   $\begin{array}{r} 30 \\ +\ 45 \\ \hline \end{array}$ = (a) 75 (b) 75 (c) 70

10. $\begin{array}{r} 348 \\ +\ 456 \\ \hline \end{array}$ = (a) 794 (b) 749 (c) 804

11. 340 + 47 + 8 = (a) 387 (b) 55 (c) 395

12. Add Sixty - Eight and Seventeen

(a) 80 (b) 85 (c) 58

13. Write down the number in the box

$\square + 14 = 71$

(a) 57 (b) 85 (c) 147_1

14. $200 + \square + 5 = 245$

(a) 35 (b) 40 (c) 205

15. Simplify (345 + 5) + (200 + 35)

(a) 785 (b) 758 (c) 350

16. $\begin{array}{r} 47 \\ -\ 19 \\ \hline \end{array}$ = (a) 82 (b) 66 (c) 28

17. $\begin{array}{r} 456 \\ -\ 348 \\ \hline \end{array}$ = (a) 108 (b) 340 (c) 804

18. Take away 95 from 205

(a) 110 (b) 300 (c) 10_1

19. $\begin{array}{r} 26 \\ \times\ 4 \\ \hline \end{array}$ = (a) 30 (b) 104 (c) 114

20. Simplify (4 x 6) + (8 x 4)

(a) 24 (b) 56 (c) 32

21. $\begin{array}{ccc} H & T & U \\ 4 & 2 & 3 \\ & \times & 4 \\ \hline \end{array}$ = (a) 419 (b) 1692 (c) 427

22. 25 X 100 = (a) 250 (b) 250 (c) 25

23. Divid 42 by 7

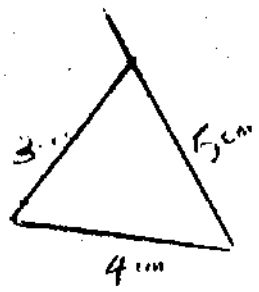
(a) 6 (b) 7 (c) 49

24. 63 ÷ 7 = (a) 9 (b) 7 (c) 70

25. Find the remainder 6 39

(a) 33 (b) 45 (c) 3

26. $\begin{array}{r} 6.12 \\ -\ 1.34 \\ \hline \end{array}$ = (a) N5.34 (b) N4.78 (c) N7.46

27. 

Find the perimeter of the triangle

(a) 8cm (b) 12cm (c) 9cm

28. How many minutes are there in 2½hours

(a) 120min (b) 150min (c) 155min.

29. How many faces has a cuboid

(a) 4 (b) 8 (c) 6.

30. Kg Kg

3.55
− 1.67 = (a) 1.88kg (b) 8.18kg (c) 4.12kg.

APPENDIX C

FIRST TRY-OUT ITEM ANALYSIS DIFFICULTY LEVEL AND DISCRIMINATIVE INDEX

S/NO	P_1	D_1	P_2	D_2	P_3	D_3
1.	.14	.31	.67x	.25	.66x	.37
2.	.70	.56	.51x	.63	.61x	.57
3.	.51x	.75	.42x	.53	.37	.46
4.	.42x	.53	.25	.44	.55x	.66
5.	.36	.47	.55x	.56	.32	.14
6.	.51x	.25	.28	.32	.35x	.57
7.	.38	.13	.39	.59	.59x	.80
8.	.28	.31	.38x	.69	.28x	.14
9.	.36	.50	.42x	.69	.53x	.71
10.	.41x	.70	.36	.81	.52x	.74
11.	.40x	.63	.39	.59	.41x	.69
12.	.36	.60	.49x	.66	.43x	.86
13.	.29	.31	.46x	.66	.43x	.68
14.	.46x	.70	.43x	.81	.37	.71
15.	.46x	.78	.45x	.78	.31	.74
16.	.45x	.65	.35	.66	.39x	.72
17.	.32	.66	.36x	.50	.42x	.70
18.	.30x	.47	.28	.09	.43x	.85
19.	.28	.56	.49x	.72	.41x	.59
20.	.38	.56	.46x	.50	.42x	.74
21.	.35x	.72	.32	.65	.44x	.71
22.	.37x	.59	.42x	.53	.20	.57
23.	.36	.66	.40x	.72	.38x	.91
24.	.33	.50	.47x	.44	.47x	.74
25.	.39x	.69	.44x	.91	.37	.86
26.	.39x	.66	.46x	.78	.36	.57
27.	.39	.35	.45x	.69	.46x	.91
28.	.21	.53	.42x	.84	.31x	.66
29.	.35x	.59	.46x	.75	.22	.48
30.	.36x	.77	.32	.66	.38x	.94

KEY

P_1 = Difficulty level of test I

D_1 = Discriminative index of test I

P_2 = Difficulty level of test II

D_2 = Discriminative index of test II

P_3 = Difficulty level of test III

D_3 = Discriminative index of test III

PRIMARY 3 MATHEMATICS FINAL TEST I

INSTRUCTIONS: For each question, three answer are provided. Find which is the correct answer. Tick the correct letter in the answer sheet.

Time Allowed: 45 Minutes.

1. Write in words: 90

   (a) Ninety

   (b) Nine

   (c) Nineteen

   (d) Nine hundred.

2. Write in Arabic Numeral: XC

   (a) 80

   (b) 90

   (c) 110

   (d) 300

3. What number does the (3) stands for in the 738

   (a) 2

   (b) 10

   (c) 30

   (d) 300.

4. What fraction of the rectangel is shaded.



   (a) 1/7

   (b) 1/6

   (c) 1/5

   (d) 1

5.
$$\begin{array}{r} 349 \\ +\ 406 \\ \hline \end{array}$$

   (a) 745

   (b) 755

   (c) 756

   (d) 57.

6. Add the following:- 27, 32, and 57

(a) 106

(b) 89

(c) 59

(d) 116

7. Subtract $\begin{array}{r} 47 \\ -\ 29 \\ \hline \end{array}$

(a) 18

(b) 75

(c) 15

(d) 28

8. Subtract $\begin{array}{r} 345 \\ -\ 232 \\ \hline \end{array}$

(a) 103

(b) 113

(c) 577

(d) 349.

9. 63 - 49 =

(a) 19

(b) 14

(c) 112

(d) 24.

10. Multiply $\begin{array}{r} 105 \\ \times\ 15 \\ \hline \end{array}$

(a) 1545

(b) 525

(c) 705

(d) 90

11. Multiply 10 x 3 x 4 =

(a) 30

(b) 22

(c) 120

(d) 40

12. $4\overline{)24}$

(a) 8

(b) 6

(c) 12

(d) 96

13. What is the time



(a) Half past five

(b) A quarter past five

(c) Three oclock

(d) Twenty minutes past three.

14. N
2.64
+ 5.65

(a) N8.29

(b) N3.01

(c) N4.56

(d) N7.29

15. How many faces has a cube?

(a) 8

(b) 9

(c) 6

(d) 12

16. Write in figures: Twenty

(a) 10

(b) 22

(c) 30

(d) 20.

17. Change to Naira and Kobo: 309K

(a) N309

(b) N30.90

(c) N39.10

(d) N3.90

Write 18 4/5 of 40

(a) 30

(b) 40

(c) 32

(d) 50

19. 58 + 69 =

(a) 28

(b) 127

(c) 117

(d) 11.

20. Write down the number in the box ☐ + 109 = 118

(a) 9.

(b) 12.

(c) 19

(d) 227.

21. Add the following: 1/5 + 2/5

(a) 2/5

(b) 2/25

(c) 3

(d) 3/5

22. Simplify (400 - 350) + (250 + 50) =

(a) 300

(b) 50

(c) 350

(d) 450

23. How many grammes are there in 2½Kg?

(a) 200

(b) 250

(c) 2.5

(d) 25.

24. Multiply 23 by 4

(a) 48

(b) 61

(c) 27

25. M CM
2 39
+ 1 72

(a) M CM
4 11

(b) M CM
5 11

(c) M CM
3 30

(d) 1M 72CM

26. 30 ÷ 3 =

(a) 3
(b) 1
(c) 112
(d) 10

27. Find the remainder 17 ÷ 3

(a) 2
(b) 1
(c) 3
(d) 5

28. Write in centimetres: 2M 32CM

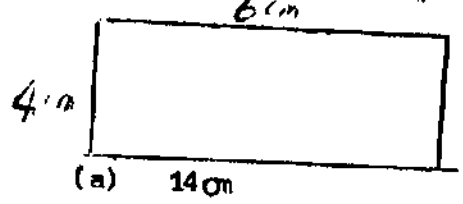
(a) 223CM
(b) 232CM
(c) 100CM
(d) 2032CM.

29. Write down the name of the figure below.



(a) Triangle
(b) Square
(c) Rectangle
(d) Cube.

30. Find the perimeter of the figure below.



(a) 14cm
(b) 20CM
(c) 18CM
(d) 10CM.

INSTRUCTIONS: For each question, three answers are provided. Find which is the correct answer. Tick the correct letter in the answer sheet.

Time Allowed: 45 Minutes.

1. XXXIV =

   (a) 24

   (b) 35

   (c) 36

   (d) 34

2. Write in figures: Eighty-eight

   (a) 18

   (b) 88

   (c) 80

   (d) 818.

3. Write in words: 22

   (a) Twenty

   (b) Twelve

   (c) Twenty-two

   (d) Twenty-one.

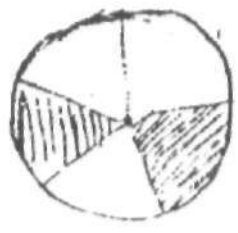
4. What fraction of the circle is shaded?

   (a) 2/5

   (b) 1/5

   (c) 4/5

   (d) 3/5



5. What fraction is 6 of 24

   (a) 1/3

   (b) ¼

   (c) 2/3

   (d) 4

6. Add $\frac{1}{6} + \frac{5}{6}$

   (a) $\frac{4}{6}$

   (b) 1

   (c) $\frac{6}{12}$

   (d) 1

7. Subtract: $\frac{1}{2} - \frac{1}{4}$

(a) ¼

(b) ½

(c) 2/3

(d) 1/8

8. Add: 349 + 277

(a) 567

(b) 626

(c) 516

(d) 72.

9. Add: 340 + 47 + 8

(a) 387

(b) 55

(c) 348

(d) 395

10. Simplify (16 x 24) + (29 – 19)

(a) 50

(b) 10

(c) 40

(d) 56

11. 978 – 609 =

(a) 369

(b) 269

(c) 279

(d) 279.

12. Add ₦5.47 + ₦3.44

(a) ₦8.81

(b) ₦8.91

(c) ₦8.03

(d) ₦8.91

13. If the cost of one hen is 49K. What is the cost of 3 hens

(a) 147K

(b) 49K

(c) 52K

(d) 98K

14. Add: Sixty-eight and Sixteen

(a) 74

(b) 84

(c) 94

(d) 52

15. Write down the number in the box ☐ + 14 = 71

(a) 1417

(b) 85

(c) 63

(d) 57.

16. Take away 95 from 205

(a) 110

(b) 300

(c) 101

(d) 22

17. Multiply:- 26
x 4

(a) 30

(b) 104

(c) 56

(d) 22.

18. Simplify (4 x 6) + (8 x 4)

(a) 24

(b) 32

(c) 56

(d) 64.

19. Multiply:

H	T	U
4	2	3
x		4

(a) 419

(b) 427

(c) 1692

(d) 1682.

20. Divide 42 by 7

(a) 35

(b) 7

(c) 49

(d) 6.

21. 6.25 ÷ 5 =

(a) 1.52

(b) 0.025

(c) 1.05

(d) 1.25.

22. Find the remainder

6) 39

(a) 33

(b) 45

(c) 3

(d) 6

23. Subtract: 456
− 348

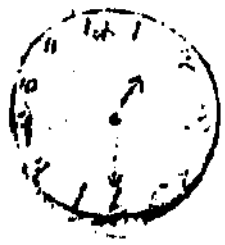
(a) 108

(b) 340

(c) 804

(d) 118

24. What time is it?



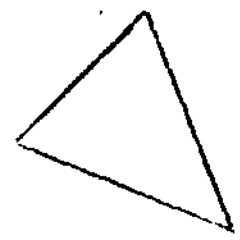
(a) 20 minutes past one

(b) Half past one

(c) One O'clock

(d) Half past two.

25. Find the perimeter of the triangle below



(a) 8cm

(b) 9cm

(c) 7cm

(d) 12cm

26. How many faces has a cuboid?

(a) 4

(b) 8

(c) 12

(d) 6

27. $\frac{1}{5}$ = $\frac{5}{\square}$

(a) 10

(b) 25

(c) 20

(d) 5

28. Write down the odd number in the set of numbers below

1, 2, 3, 5, 6, 7, 8, 9, 10, 11

(a) 1, 2, 3, 4, 5,

(b) 3, 5, 6, 9, 10

(c) 3, 6, 7, 9, 10

(d) 3, 5, 7, 9, 11

29.
Kg.
3.55
− 1.67

(a) 1.88kg

(b) 8.18kg

(c) 4.12kg

(d) 2.88kg

30. How many centimetres are there in 1m 49cm

(a) 149cm

(b) 49cm

(c) 100cm

(d) 1049cm

LISSAFIN AJI JARABAWAR KAKSHE KASHI NA FARKO

BAYANI: A ko waco tambaya akwai amsoshi guda hudu. Ka nemi amsa guda wadda take daidai sannanka sanya amsarka a cikin akwatin amsoshi.

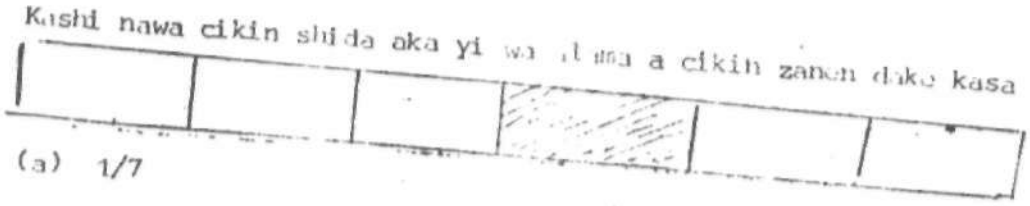
LOKACI: Minti 45

1. Rubuta cikin Kalmomi: 90
   (a) Casain
   (b) Tara
   (c) Goma sha tara
   (d) Dari tara

2. Mene ne ma'anan wannan:- XC
   (a) 80
   (b) 90
   (c) 110
   (d) 1010.

3. Mene ne matsayin 3 a cikin 738
   (a) 3
   (b) 10
   (c) 30
   (d) 300.

4. Kashi nawa cikin shida aka yi wa alama a cikin zanen dake kasa



   (a) 1/7
   (b) 1/6
   (c) 1/5
   (d) 1

5. 406
   + 349
   (a) 755
   (b) 745
   (c) 756
   (d) 57.

6. Tara: 1/6 + 5/6

(a) $\frac{4}{6}$

(b) 6/12

(c) 5/36

(d) 1

7. Debe: $\frac{1}{2} - \frac{1}{4}$

(a) 1/4

(b) 1/2

(c) 2/3

(d) 1/8

8. Tara: 349
+277

(a) 567

(b) 626

(c) 516

(d) 72.

9. Tara: 340 + 47 + 8

(a) 387

(b) 55

(c) 395

(d) 348

10. Gano: (16 + 24) + (29 - 19)

(a) 50

(b) 10

(c) 40

(d) 56

11. 978 - 609 =

(a) 369

(b) 269

(c) 297

(d) 279

12. Tara: N 5.47
+ N 3.44

(a) N2.03

(b) N8.81

(c) N8.03

(d) N8.91

13. Idan ana sayar da kaza daya akan kwabo 49, nawa ne kudin kaji biyu?

(a) 147K

(b) 98K

(c) 52K

(d) 49K

14. Sittin da takwas a tara da Gama sha shida

(a) 74

(b) 84

(c) 94

(d) 52.

15. Rubuta alkalumman da ya dace a cikin wannan akwati [ ] + 14 = 71

(a) 1417

(b) 85

(c) 63

(d) 57.

16. Debe 95 daga cikin 205

(a) 110

(b) 300

(c) 101

(d) 115

17. Nemo amsar wannan: 26
x 4

(a) 30

(b) 104

(c) 56

(d) 22

18. Gano: (4 x 6) + (8 x 5)

(a) 24

(b) 32

(c) 56

(d) 64.

19. 58 + 69 =

(a) 28

(b) 127

(c) 117

(d) 11.

20. Nemo alkalumma da ya kamata a sanya cikin akwati

___ __ + 109 = 118

(a) 9

(b) 17

(c) 19

(d) 227.

21. Tara waddan nan 1/5 + 2/5

(a) 2/5

(b) 2/25

(c) 3

(d) 3/5

22. Gano: (400 - 350) + (250 + 50)

(a) 300

(b) 50

(c) 350

(d) 450

23. Giram nawa ke cikin kilo $2\frac{1}{2}$?

(a) 200

(b) 250

(c) 2.5

(d) 25.

24. Mene ne 23 X 4

(a) 48

(b) 61

(c) 27

(d) 92

25. M CM
2 39
+ 1 72

(a) 4m 11cm

(b) 5m 11cm

(c) 3m 30cm

(d) 1m 72cm.

26. 30 ÷ 3 =

(a) 3

(b) 1/10

(c) 33

(d) 10.

27. Me ya ragu in an raba 17 gida 3

(a) 2

(b) 1

(c) 3

(d) 5.

28. Rubuta cikin santimita: 2m 32cm

(a) 223cm

(b) 232cm

(c) 100cm

(d) 2032cm

29. Menene sunan wannan zane?



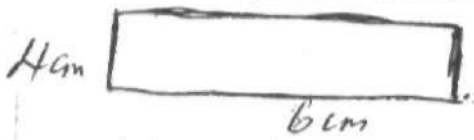
(a) Zane mai kusurwa uku

(b) Zane mai tsawo da fadi daidai

(c) Zane mai tsawo da fadi bamban

(d) Zane mai tsawo daidai.

30. Nawa ne gewayen wannan akwatin na kasa



(a) 14cm

(b) 20cm

(c) 18cm

(d) 10cm

LISSAFIN AJI UKU JARABAWAR KARSHE KASHI NA BIYU

BAYANI: A ko wace tambaya akwai amsoshi guda hudu, Ka nemi amsa wadda take daidai, Sannan ka sanya amsarka cikin akwatin amsoshi.

LOKACI: Minti 45

1. XXXIV =
   (a) 24
   (b) 35
   (c) 36
   (d) 34

2. Rubuta cikin alkalumma: Tamanin da takwas.
   (a) 18
   (b) 88
   (c) 80
   (d) 818

3. Rubuta cikin kalmomi: 22
   (a) Ashirin
   (b) Goma sha biyu
   (c) Ashirin da biyu
   (d) Ashirin da daya.

4. Kashi nawa aka yi wa alama a zanen da ke kasa
   (a) 2/5
   (b) 1/5
   (c) 4/5
   (d) 3/5



5. Mene ne kashi 6 daga cikin 24
   (a) 1/3
   (b) 1/4
   (c) 2/3
   (d) 4/4.

6. Tara: 1/6 + 5/6

(a) $\frac{4}{6}$

(b) 6/12

(c) 5/36

(d) 1

7. Debe: $\frac{1}{2} - \frac{1}{4}$

(a) 1/4

(b) 1/2

(c) 2/3

(d) 1/8

8. Tara: 349
+277

(a) 567

(b) 626

(c) 516

(d) 72.

9. Tara: 340 + 47 + 8

(a) 387

(b) 55

(c) 395

(d) 348

10. Gano: (16 + 24) + (29 - 19)

(a) 50

(b) 10

(c) 40

(d) 56

11. 978 - 609 =

(a) 369

(b) 269

(c) 297

(d) 279

12. Tara: M 5.47
+ M 3.44

(a) M2.03

(b) M8.81

(c) M8.03

(d) M8.91

13. Idan ana sayar da kaza daya akan kwabo 49, nawa ne kudin kaji biyu?

(a) 147K

(b) 98K

(c) 52K

(d) 49K

14. Sittin da takwas a tara da Goma sha shida

(a) 74

(b) 84

(c) 94

(d) 52.

15. Rubuta alkalumman da ya dace a cikin wannan akwati ☐ 14 = 71

(a) 1417

(b) 85

(c) 63

(d) 57.

16. Debe 95 daga cikin 205

(a) 110

(b) 300

(c) 101

(d) 115

17. Nemo amsar wannan: 26
x 4

(a) 30

(b) 104

(c) 56

(d) 22

18. Gano: (4 x 6) + (8 x 1)

(a) 24

(b) 32

(c) 56

(d) 64.

19. 58 + 69 =

(a) 28

(b) 127

(c) 117

(d) 11.

20. Nomo alkalumma da ya kamata a sanya cikin akwati

____ + 109 = 118

(a) 9

(b) 17

(c) 19

(d) 227.

21. Tara waddan nan 1/5 + 2/5

(a) 2/5

(b) 2/25

(c) 3

(d) 3/5

22. Gano: (400 - 350) + (25 + 50)

(a) 300

(b) 50

(c) 350

(d) 450

23. Giram nawa ke cikin kilo 2½?

(a) 200

(b) 250

(c) 2.5

(d) 25.

24. Mene ne 23 X 4

(a) 48

(b) 61

(c) 27

(d) 92

25. Nawa ne tsawon wannan zane da ke kasa?



(a) 8cm
(b) 9cm
(c) 7cm
(d) 12cm

26. Fuskoki nawa akwati ke da su:

(a) 4
(b) 8
(c) 12
(d) 6

27. $\frac{1}{5} = \frac{5}{\square}$

(a) 10
(b) 25
(c) 20
(d) 5

28. Nemo alkalumman da biyu bata iya shiga cikinsu ba tsate da ragowa ba?

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, ..

(a) 1, 2, 3, 4, 5,
(b) 3, 5, 6, 9, 10
(c) 3, 6, 7, 9, 10
(d) 3, 5, 7, 9, 11

29.
Kg
3.55
- 1.67

(a) 1.88kg
(b) 2.18kg
(c) 4.12kg
(d) 2.88kg

30. Santimita nawa ke cikin 1M 49CM

(a) 149cm
(b) 49cm
(c) 100cm
(d) 1049cm

APPENDIX F

DIFFICULTY INDEX AND DISCRIMINATIVE FOR THE FINAL TEST

S/NO	P_1	D_1	P_2	D_2
1.	0.87	0.42	0.78	0.33
2.	0.53	0.18	0.74	0.34
3.	0.56	0.24	0.62	0.85
4.	0.52	0.51	0.61	0.51
5.	0.61	0.68	0.52	0.60
6.	0.58	0.52	0.43	0.66
7.	0.62	0.55	0.48	0.34
8.	0.66	0.58	0.66	0.58
9.	0.62	0.78	0.52	0.56
10.	0.51	0.42	0.48	0.42
11.	0.54	0.87	0.58	0.45
12.	0.62	0.76	0.48	0.66
13.	0.51	0.42	0.37	0.54
14.	0.58	0.75	0.53	0.76
15.	0.45	0.54	0.53	0.79
16.	0.50	0.66	0.48	0.41
17.	0.48	0.64	0.56	0.62
18.	0.50	0.51	0.56	0.59
19.	0.57	0.69	0.53	0.71
20.	0.53	0.69	0.42	0.27
21.	0.47	0.43	0.45	0.35
22.	0.57	0.42	0.46	0.52
23.	0.50	0.46	0.52	0.64
24.	0.52	0.59	0.49	0.56
25.	0.53	0.61	0.48	0.69
26.	0.54	0.57	0.45	0.77
27.	0.52	0.64	0.49	0.56
28.	0.50	0.46	0.46	0.72
29.	0.55	0.60	0.53	0.46
30.	0.42	0.59	0.49	0.74.