DEVELOPMENT AND STANDARDISATION OF EARLY READING LITERACY TEST FOR PRE-SCHOOL CHILDREN IN OYO STATE, NIGERIA

BY

OLUKEMI OLUREMI, ADEOSUN

MATRIC. NO.: 79996
B.Ed (Ibadan), M.Ed (Ibadan)

A Thesis in the International Centre for Educational Evaluation (ICEE)
Submitted to the Institute of Education
In partial fulfilment of the requirement for the degree of

DOCTOR OF PHILOSOPHY

of the

UNIVERSITY OF IBADAN

AUGUST, 2019

ABSTRACT

Standardised tests are capable of measuring pre-school children's latent reading ability and are also used to assess examinees' performance across norms (gender, age, school-type and location). However, in Nigeria and Oyo State in particular, most of the available tests are foreign and not adapted to the socio-cultural learning environment. Consequently, early reading ability of pre-school children is not uniformly assessed. Previous studies have focused mainly on primary and secondary school levels with little emphasis on standardised reading literacy test at the pre-school level. This study, therefore, developed and standardised Early Reading Literacy Test (ERLT) for pre-school children along the norms with the aim of deriving valid and uniform instrument to assess the reading ability of pre-school children in Oyo State.

The study was anchored to Cognitive Theory of Literacy Development, while Instrumentation Approach and counter-balance design were adopted. The study was conducted in two stages. In stage one, the ERLT (with identification of letters; objects; colours; signs and symbol; reading fluency and picture reading subsets) was developed and validated (r=0.84) using a sample of 45 pre-school children from two randomly selected pre-schools in Osun State. For instrument standardisation, the proportionate-to-size sampling technique was used to select four Local Government Areas (LGAs) each from Oyo Central and Oyo North senatorial districts, and three LGAs from Oyo South senatorial district. Three percent of public (21) and private (34) pre-schools were selected across the 11 LGAs, while 55 intact classes (one per school) with 776 pre-school children were sampled. Data were analysed using T-score and Stannine.

The discrimination indices of DRAFT-ERLT ranged between 0.12 and 6.36, while the difficulty parameter was between -7.92 and 0.83, and a total of 163 out of 226 items were finally assembled and administered. The ability test scores ranged from 19.80 to 71.70 ($\bar{x}$=51.0). Normative scores established using stannine 1, 2 and 3 reflecting 23.0% of examinees were below average while 4, 5 and 6 showing 54.0% of the examinees were on the average. Stannine 7, 8 and 9 showing 23.0% of the examinees were above average. The pre-school children mean scores by age were below four years ($\bar{x}$=57.85}; four years ($\bar{x}$=50.86); five years ($\bar{x}$=50.85); gender were male ($\bar{x}$=51.79); female ($\bar{x}$ =50.40); school-type were private ($\bar{x}$=53.58); public ($\bar{x}$=47.73) and school location were urban ($\bar{x}$=54.28); rural ($\bar{x}$=41.61). Differential Item Functioning was observed among groups (gender, school- type and location) in the identification of some letters: w, m, b, d, v, x, q, s and z; identification of objects like bag, tree, egg, tyre and fan; and picture reading on a boy fetching water and a girl sweeping the floor.

The developed and standardised early reading literacy test was sensitive only to gender, location and school-type. Pre-school teachers should ensure gender balancing and devote more time to teaching letters w, m, b, d, v, s and x, particularly for pre-school children in public schools and rural areas.

Keywords: Nigeria pre-school children, Early Reading Literacy Test, Differential Item Functioning.

Word count: 478

DEDICATION

This thesis is dedicated to the Almighty God the Alpha and Omega of my life, and to my mother, Mrs, Olufunke Dorcas Bello.

ACKNOWLEDGEMENTS

All glory, adoration and honour be unto the Ancient of Days, the one that promises and brings to pass, though it may tarry, it will surely be fulfilled. Thank you Jesus for seeing me through this study.

I am highly grateful to my supervisor, Dr. Monica N. Odinko for her mentoring, patience and sacrificial efforts in ensuring the successful completion of this work. Thank you ma for your motherly support. I also thank all my lecturers in the Institute of Education, University of Ibadan; Professors Yoloye, T. W., E. Adenike Emeke, I. Isiugo-Abanihe, Folajogun V. Falaye, Adewale, J. G. Adegbile, J. A., Onuka, A., E. A. Okwilagwe and Drs M. M. Osokoya, Ibode, F. O., Adeleke, J. O., S. F. Akorede, Ikmat Junaid, Abijo, J. A., Babatunde, E., Metibenu, M. A., Olutayo, T. Omole, Oladele, B. K. and Desmenu, O. M. for their support, love and patience during the course of this study and the process of empowering me with the basic rudiments in research. Special thanks to Professor Adegbile, J. A. for standing as my father and counsellor even outside academics. I cannot forget Prof. Kester, K. O. for equipping me with his professional expertise in abstract writing. Thank you sir for being a good model of a selfless life.

I sincerely appreciate my mother Mrs. Dorcas Olufunke Bello and my late father, Prince Bello Olasupo for their parental love and opportunity given to me to have formal education. My siblings – Bola Akamo and Sina Bello, I appreciate your love, support, prayers and encouragements when it seemed there was no way out. Elder and mummy Michael Bakare for your spiritual support in ensuring the dream is actualised, Dr. Bello, B.V. for ensuring that I enjoyed a good medical attention and guidance in the course of this study. Drs. Dipeolu, F., Raheem, A. W., Adesina, O. J. for given me a solid foundation in research even from the undergraduate level. Pastor Olajire Oluseyi, Mrs. Folakemi J. Ayodiran, Mojirayo F. Okeyinka, A. Olaniran, Helen O. Fadipe, Helen Osun, Ganiyat Akinyemi, Prince Adelabu, A. O., Mr. Ajao, A. O., Babatunde, W. A., Mukaila Azeez, Mrs. B. A. Aremu and all the members of staff of Oyo State Universal Basic Education Board.

I am also indebted to all my colleagues in the Institute of Education, University of Ibadan. Dr Morinola Fakayode, Dr Enu, Valentine O., Mr. Shola Ajala, Mrs. Shade Oluwole, Dr S. Alieme, Mrs. Benedicta Oyeme, Dr Deborah Oyegoke, Dr Nathaniel Olaniran, Dr Alfred Adeniyi, Mrs. Hafsat Akanni, Mrs. Nkem Adewoyin, Dr Abdul Afeez Jinadu, Mrs. Abdul-Azeez Fatai Ayiki, Dr Ayanwale Musa, Dr Atanda Abiodun, Dr Motunrayo K. Oyaremi, Mrs. Toyiin Rasak and Mr. Godwin

Atsua. I cannot mention all the names, but I thank you all from the depth of my heart for given me all the supports I needed in the course of this study. My gratitude also goes to my husband, father and pastor, Adeosun Michael Ayodele and my lovely children – Mr. Adeosun Ayanfeoluewa A. and Adeosun Oreofeoluwa F. for your understanding, supports, encouragements and prayers throughout the years of this study. Thank you all for loving me. Finally, I am grateful to all that have assisted me at one level or the other in the completion of this study. The all sufficient God will be your sufficiency at all times in Jesus name.

CERTIFICATION

I certify that this study was carried out by Olukemi Oluremi ADEOSUN (Matriculation Number: 79996) under my supervision at the International Centre for Educational Evaluation (ICEE), Institute of Education, University of Ibadan, Ibadan, Nigeria.

…….. ……

Dr. Monica Ngozi Odinko **Date**

B.Ed(Ibadan), M.Ed ,(Ibadan), Ph.D (Ibadan), Ph.D (Edinburg)
Reader, International Centre for Educational Evaluation
Institute of Education,
University of Ibadan, Ibadan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE
INTRODUCTION

1.1 Background to the Problem

The physical, psycho-social, emotional, cognitive and language development of a child start from the conception.The type of nurture given to children from birth to five years may either make or mar their development. Parents and other stakeholders may therefore, need educationalprogrammes that may enhance holistic development of children. Two of these educational programmes are Early Childhood Care and Development Education (ECCDE) and Pre-Primary Education.

The Early Childhood Care and Development Education (ECCDE) is the care, protection, stimulation and learning promotion of children from birth to four years in crèche or nursery, while Pre-Primary Education is the one year education given to children aged 5 prior to their entering primary school (Federal Republic of Nigeria, 2014). Studies like National Association for the Education of Young Children (1998) and Vidya (2014) have shown that attending ECCDE/Pre-Primary schools does not only increase children's readiness for formal schooling, but also enhances positive long-term improvement in participants' school performance and social outcomes. The ECCDE aims at ensuring a smooth transition from home to school and preparing children for future learning activities. Efforts to support and provide quality ECCDE will promote growth in cognitive, language acquisition, motor skills acquisition and other adaptive skills acquisition to enable children function well during other levels of education (Oduolowu, 2011).Thus, ensuring that quality and standard foundation are provided for children is important since the quality of nurture given to them during early years is crucial to their all-round development.

Moreover, Early Childhood Care and Development Education is an integral part of basic education (Federal Republic of Nigeria, 2014) which represents the first and essential step towards achieving the Mellenium Development Goal 2 by 2015 (Federal Republic of Nigeria, 2005). The targeted year, 2015, had come and gone, yet, the MDG 2 goal has not been fully realised in Nigeria and if Sustainable Development Goal 4.2 (2030) would be achieved, there is a need for increasing efforts towards promoting quality Early Childhood Care and Development and Pre-Primary Education programmes. Oduolowu (2011) documented the benefits of ECCDE to include better critical thinking and attention skills, better reading and writing skills, ability to handle the demand of formal schooling, higher graduation

rates and preparation of children to succeed in life. The Universal Basic Education Commission (UBEC) thus, views the need to place Pre-Primary education at the centre of education policy and in response pays more attention to early reading literacy considering its importance in enhancing academic achievement and nation's development. It thus becomes necessary for stakeholders to harness all measures to ensure that early reading literacy skills are assessed beginning from the early years. Reading is the process of constructing meaning from a written text (Sonnen-Schein, Stepleton and Benson, 2010). Ijaiya (2007) also described reading as a complex interaction between the text and the readers based on the readers' previous knowledge and experience. In reading, an individual acquires language and communicates by sharing information and ideas with others. It is therefore, appears that a child's daily experiences, interactions and communication with others will go a long way in shaping his or her reading skills. Thus, effective early reading tasks enable children to become fluent readers who comprehend what they read, communicate and apply their knowledge and skills in new contexts (Wanda, 2013). A child who reads well is more likely to perform better in whatever he or she learns because the child's spelling, vocabulary and comprehension skills are most likely developed.

Although, structured formal instruction in reading does not begin at pre-primary school, children develop early reading skills through engaging and seeing others engaged in print, talk with others, listen and tell stories (Roskos and Vukelich, 2006). Through children's interaction in the environment, they develop oral language skills, phonological and phonemic awareness, knowledge of what print is, how it is used as well as knowledge of story structure (Ijaiya, 2007; Kaderavek and Pertimonten, 2014). Observation, in real life situation, shows that in their day- to- day activities, children observe and listen to adults as they speak and read. They come in contact with prints like story books, newspapers, magazines, textbooks, signs, charts and calendars. All these could assist in stimulating major reading sub-skills in children.

Kaderavek and Pertimonten (2014) claimed that children acquire most of what they know about oral language by listening and speaking with others, including families, peers and teachers. In the process, they build vocabulary, semantic (awareness of meaning) and syntactic knowledge (awareness of structure) that form the foundation for reading and writing skills. Children who are proficient in oral language have a solid beginning for reading; they identify words accurately and

predict what the written language means (Roskos and Vukelich, 2006). Through real life observation, children develop these skills better when adults allow them to express themselves, ask questions and relate their experiences with their peers and other adults. A child's oral language ability can be tested by asking the child to identify familiar objects around them and read short sentence like two to four letter words with familiar objects.

Phonological awareness is another important determinant of success in learning to spell and read. It involves the detection and manipulation of sounds at three levels of sound structure; syllables, onsets and rhyme and phonemes. Awareness of these is demonstrated through a variety of tasks such as discrimination, syllable identification and sound awareness (Gillon, 2017). The researcher further stated that for most children, strong readers have strong phonological awareness skills, while poor readers have poor phonological awareness skills and that acquiring these skills (such as stress, accent, intonation, syllable structure and sounds that are distinctive units within a language) in pre-school years can predict how well a child will read in the school years ahead. Measuring a child's ability to identify letters, demonstrate alphabetic sounds in English Language, sing nursery songs and rhymes are some of the likely ways of finding out the child's phonological awareness level.

Knowledge about print according to Hood, Conlon and Andrews (2008) is another area necessary in developing early reading literacy skills. When children first encounter prints, they may not be aware that symbols on the page represent spoken language or that they can convey meaning. Knowledge about print involves directionality (for example, knowing that English text is read from left to right and top to bottom), differences between letters and words, awareness of capitalisation and punctuation, as well as common characterisation of books such as the front and back, title and author. Young children can be taught these concepts by interacting with and observing experienced readers including teachers and family members who draw their attention to prints and give them opportunities to demonstrate their understanding of the concept.

Reading fluency and comprehension are other aspects of reading skills acquisition which come through practicing reading books that contain primarily familiar subjects. Reading fluency is the ability to read with expression (Wagel, Marlin and Benneth, 2006). Shanahan and Lonigan (2013) stated that fluent reading can be modeled by reading aloud to children by using expression and acting the

reading. Words that pre- school children hear on daily basis in their environment and objects that they are familiar with can be combined to form sentences, starting from very short ones like two- letter word to four- letter word. These may be used to test their reading fluency as observed through interaction with children. As children develop reading fluency skills, they improve on their ability to read more extensively with proper phrasing; thereby gaining more of the text's meaning (Tracey and Morrow, 2006). Story comprehension is a way of appropriating meaning from text and picture. A reader must translate print to language and interpret its meaning. Reading story books to young children will familiarise them with story structure and in turn, help to facilitate comprehension (Tracey and Morrow, 2006). Discussing the text may familiarise the children with strategies needed to comprehend such text.

Chall (1993) explained that all individuals progress through stages of reading acquisition as highlighted below:

Stage 1: Preliminary reading (birth to age 6);

Stage 2: Early reading or deciphering (ages 6-7);

Stage 3: Validation, eloquence and identification of printed objects (ages 7-8),

Stage 4: Learning the new concept (ages 8-14);

Stage 5: Manifoldopinion prints (ages 14-18); and

Stage 6: Construction and deconstruction (age 18 and above). Progression through the stages is characterised by the recognition and decoding of words by learning the meaning of uncommon words (abstract words, ideas and concepts).

Over the years, attention has been paid to understanding differences in reading literacy ability of children across cultures. Many of these studies focused on the relationship between the children demographic variables (e.g. age, gender, school type and location) and reading literacy abilities. This study also looked at children performance in early reading literacy vis a vis their age, gender, type of school they attended and its location. Research findings confirmed that boys and girls have differential abilities. For example, Lynn (1994) and Ijaiya (2007) found that female performed significantly better on word fluency test, while male achieved significantly higher score in spatial test. Ijaiya (2007) added that in general, girls speak earlier than boys and excel in word usage, correctness of sentence structure and comprehensibility of speech more than boys. School type is another important variable that predicts performance in reading literacy. A school with literacy-rich-environments where pre-school children have access to picture books, big books and other print materials

contribute positively to child literacy and language development (Hartas, 2011). Other important parameters embedded in school type factors include teachers' teaching experience, time devoted for reading activities, method and inter-personal relationship between teacher and learner (Ijaiya, 2007; Sonnen-Schein, Stapleton and Benson, 2010).

The differences in academic performance due to location could be as a result of preference by teachers to work in some locations than others. Johnson (2011) concluded that highly qualified teacher prefer to serve in the urban areas than in the rural areas. Many teachers do not accept posting to rural areas and even if they accept the posting, they do not live in those rural areas, this does not give room for total commitment to their duties. This invariably affects the performance of students. Odinko (2002) researched on some home and school factors determining literacy skills and found that school location also influences school income, urban schools that earn more income are more likely to provide books, equipment and other instructional materials that would enhance effective learning in such schools, whereas, schools in rural area rarely provide these materials. Ofodu and Lawal (2010) carried out a study on effect of school location on learner's performance in reading comprehension and the findings indicated that school location had effects on students' performance in reading comprehension. This however, contradicts the findings of the Sara (2017) which revealed school location had no significant effect on students' academic performances.

Considering the differences in reading literacy ability of pre-school children, assessment in Pre-Primary education becomes crucial. The process of collecting information concerning children from various forms of evidence, then organising and interpreting that information is called assessment (McAfee, Leong and Bodrova, 2004). Assessment of young children helps researchers to measure the appropriateness of the experiences they are exposed to. Results emanating from assessment in Pre-Primary education are used for various purposes: to promote children learning and development, identify children for health and social services, monitor trends and evaluate programmes and services, and assess academic performance to hold individual learner, teacher and schools accountable (Bedford, Walton and Ahn, 2013; Shepard, and Smith 1988).

According to Groulund (2006), assessment can be formative or summative. Formative assessment is also referred to as assessment for learning. This is an on-

going process of seeking, gathering and interpreting evidences on an individual learner. It allows learners to pinpoint how much they know and identify areas that need improvement. Usually, where studentss are not scored, the assessment acts as a scale to their learning progress and allows re-thinking and re-delivery of instructions. Summative assessment takes place after the learning has been completed. It provides information and feedback that sums up teaching/learning processes. Summative assessment is also known as assessment of learning; it is the test of knowledge and retention on a subject matter or development stage of a child in a particular domain. Assessment as learning occurs when students actively engaged in self or peer assessment. They learn about themselves, how they learn as students and device means of improving on their learning performances.

According to Whitehouse (2002); Zigler and Valentine (1979), the Early Childhood Care and Development Education (ECCDE) projects initiated by the United States government stressed the importance of improving ECCDE programme. This led to an increased call for standards-based testing for children. Standardised test and other assessment methods are now being adopted for kindergarten to determine how children perform based on curriculum contents they are exposed to and whether they measure up with other children within their age group. Performance assessment, permits children to demonstrate what they understood through the performance of a task. It helps teacher in seeking information on the child's development and accomplishments in all domains. When performance assessment is combined with other assessments, a longitudinal record of change in a child's development is observed and should be documented. Such assessment could involve the use of teacher-made test or standardised test. A standardised test is designed to assess the abilities, knowledge, skills of individuals under clearly specified and controlled conditions relating to construction, administration, scoring and interpretation. Raw scores are normed in standardised test (Jacob, Darrel, Paulgros, 2018). This type of test requires all test takers to answer the same questions selected from the same questions bank that were developed based on the curriculum contents such learners were exposed to. These questions may be in form of objectives, true/false, short answer, prose questions andmixture of questions types.

There are many examples of standardised tests produced and made available for use in developed nations. They include Enhanced Kindergarten Performance Standards published in 1996 for children from birth to age 5, Early Head Start

published in 2000 (Street and Suite, 2000) and Peabody Picture Vocabulary Test published in 2013 (Eigsti, 2013), meant for individual within two years five months and eighteen years. Standardised test can be a performance test. Standardised performance tests are designed to; measure the knowledge and skills that learners have learnt, determine the academic progress they have made over a period of time, compare their scores through norming and identify the appropriate academic placement for a learner (Moshtaghian, 2009). Standardised performance tests are also used as measures of effectiveness. These tests are useful in evaluating whether students have learnt what are expected of them and whether they have met the state learning standards. For example, in Nigeria, Early Childhood Care and Development Education has a learning standard that covers developmental domains expected of a child. Identification of performance gaps among different test taker groups is another reason for standardised performance test. This study developed and standardised an instrument to measure early reading literacy skills at pre-school level in Nigeria and also checked whether pre-school children have met the learning standards in language development domain (early reading literacy skills).

The unique quality of standardised test is uniformity in test administration using precise administration procedures. Standardised tests are also quantifiable that is, numerical scores are translated into a derived score which is used to evaluate the child's performance and compared with the established standard. Unlike informal measurement strategies, standardised tests are reliable and valid. Children scores may be interpreted with assurance to precisely reveal each learner's behavior (Bowma and Burns, 2001).

Another importance of standardised test according to Ronald and Merk (2010) is that standardisation sample allows the test designer to create a normal distribution which can be used for comparison with any specific future test group, that is, norm group. Tests that are standardised are presented with one of the interpretive framework called norm. It can be norm–referenced or criterion – referenced. A norm referenced framework discovers the actual position of learner’s score in an array of scores. Criterion referenced is needed once performance in a test gives the extent to which the ability of a learner conforms with the standard based, criterion or ability level (Glaser, 1988). A variety of score conversions are applied to enable inter-individual comparison after norm-referenced tests are employed which comprise rank percentile, Z – score, T – score, Stannine and normal curve (Osadebe, 2014; Osadebe,

2001; Owen and Jones, 1994).

Standardised tests are normed across fairly large groups by collecting large amounts of testing information from large cohorts of test takers and later comparing performance of test takers in groups. Norm serves as a frame of reference for test score interpretation. Normed tests do not only reflect mastery of specific cognitive abilities but also allow a child's performance to be compared with other children of normative group. Age equivalent norms are specifically designed for use as a reference in the context of the age of the test takers who achieve a particular score (Ronald and Merk 2010). It is the comparison of individual's performance with others in a group (age-wisely)to know whether the individual achieved at a level above, below or is equal to average performance of age equivalent test taker. Test scores would be invalid for testing a sample that is not reflected in the normative group. In this study, Early Reading Literacy Test was normed across age, gender, location and type of school that a pre-school child attends for comparison.

American Psychological Association (APA, 1996) based the following important steps in standardising a test:

(i) define clearly the reason(s) for the test and how the test scores will be used-whether for certification, placement, licensing and or for other purposes;

(ii) define area of construct (capacity, body of knowledge and or skills to be measured);

(iii) determine the test format - how the test will be presented and how the test takers will respond;

(iv) develop the experimental forms;

(v) consult experts in the field for validation (face and content);

(vi) assemble selected experimental items for trying out and analysis;

(vii) administer the final test form on another larger sample in order to acquire norm data; and

(viii) develop the test manual which should describe the purpose of the test, the development of the test, standardisation procedures and information on validity, reliability and method used to select the norming group.

In recent times, Early Childhood Educators in Nigeria are attaching great importance to the assessment of early reading literacy. This may be due to a very poor reading literacy performance observed among primary school children in both lower and higher classes probably as a result of the weak pre-school reading literacy

foundation. In assessment, instrument must be valid and reliable; therefore, efforts should be made to ensure that the test items are reliable and actually measured what they are supposed to measure because decisions will be made based on the analysis of the assessment. In educational measurements, there are some frameworks through which tests can be developed, validated, and used for assessing student's performance. These include Classical Test Theory (CTT) and Item Response Theory (IRT). The Classical Test Theory comprises three concepts. These are; observed score, true score and error score. According to Hambleton and Jones (1993), several models have been formulated within these three concepts of which the central model is the Classical Test Model (CTM). This model links the observed test score (X) to the sum of the two unobserved (called latent) variables, true score (T) and error score (E). Mathematically, the Classical Test Model (CTM) is represented by $X = T + E$. In this equation, there are two unknown scores, the true score (T) and error (E), thereby making the equation not solvable. However, Hambleton and Jones (1993) presented that the use of Classical Test Model is made possible by three assumptions, including:

(a) True score (T) and error score (E) are uncorrelated;
(b) The average error score in the population of examinees is zero, and
(c) Error scores on parallel test are uncorrelated

Thus, in the Classical Test Theory (CTT), the examinee's test score would be the sum of the scores received on all the items in the test. This referred to as number-correct scoring (Tomkowickz and Wright, 2007). However, this method of scoring produced maximum likelihood trait estimates based on raw scores (that is, total number of correctly answered items). In this method, examinees who answered correctly the number of items irrespective of items level of difficulties and discrimination earn the same scale score. This means the nature of the items parameters (that is, difficulty and discrimination levels) are not considered in the scoring of the examinees performance. However, difficulty and discrimination indices are statistics that guide test development (Meyer and Zhu, 2013). Test developers use these indices to identify problematic items such as those that are two easy or too difficult for students that are unrelated to the overall test score (or items with low discrimination). The relationship between item statistics and test statistics such as test-score, mean, standard deviation and reliability are used in the test development process to produce tests of the desired statistical properties. For example, in the construction of norm-referenced tests, items with difficult level less than 0.2 and

greater than 0.8 are deleted. Similarly, items with discrimination level less than 0.2 are considered as bad items. Also, items which are found to be biased against or for an identifiable group of examinees are also considered poor (Crocker and Algina, 1986; Dodeen, 2004).

Classical Test Theory (CTT) has been used to estimate psychometric properties of a test and validate its use for program assessment. Furthermore, CTT procedure is concerned with the reliability of a test and assumed that the items of the test are sampled at random from a domain of relevant items, CTT analysis relies on the total score of the test takers in the measurement of the test parameters and the ability of the test taker (Dodeen, 2004). Classical test theory (CTT) analysis is still in use today but there are some problems associated with its use. According to Hambleton and Jones (1993), there are three major disadvantages of Classical Test Theory (CTT) in estimating the psychometric properties of a test. These are;

i. Assumption of equal errors of measurement among all test takers. For example, test takers with low ability might not answer correctly items with high difficulty index and yet get the same scores with other test takers.
ii. Item difficulty and item discrimination in CTT are group dependent. The estimates cannot be used to generalise but depend only on the group of the test takers from which they are estimated.
iii. CTT has no correlation of true and error scores, meaning that the scores a test taker gets depend on the group along which he/she is measured. This is fund cognitively wrong, because two estimates will emerge for a particular trait and each test taker has one trait level.

IRT came up as a recent theory developed to tackle the inadequacies of Classical Test Theory. It premise is that responses to test item differ due to individuals 'capacity or hidden trait of underlying concept measured by the test. Hambleton and Jones (1993) and Rupp (2009) explained that any theory on (IRT) put forward by Lord (1992) can be described by defining learner's individual differences denoted as traits or capabilities and by estimating scores for students concerning the hidden behaviour. Such scores are called 'ability scores' and can be used to predict or explain item and test performance. Traits that are not directly measurable are referred to as latent traits or abilities in cognitive testing and are not yet very noticeable in children. Item Response Theory is a general statistical theory about examinee, item, test performance and how performance relates to the abilities that are being measured

by items in the test (Hambleton and Jones, 1993).

Item Response Theory attempts to model the ability of a test taker and the probability of answering an item correctly based on the pattern of responses to all the items that constitute the test. Under IRT, the primary interest is in whether an examinee gets an item correctly or not rather than in the raw test scores. This is referred to as item-pattern scoring procedure (Tomkowickz and Wright, 2007). Some of the advantages of IRT over CCT are, IRT approaches are different due to the fact that they are item-centred rather than test-centreed. The measures estimate statistics for individual item in the assessments vis-à-vis the ability level of respondents. Item characteristics curve or function shows the actual connection among the unseen variable (abilities) and the observed response. Also, derived from IRT lends itself to modern test formats in which individual's response to chosen items are based on the examinees ability levels other than all items in the test.

The item-pattern scoring method produces maximum likelihood trait estimate based on pattern of item responses. Item-pattern scoring takes into consideration not only how many items the examinee answered correctly but also which items were answered correctly. As such examinees who answered the same number of items correctly may get different scale score if they answered different items correctly (Tomkowickz and Wright, 2007).To estimate the ability of the examinee from his/her response to a particular test items, the items parameters of the test should be taken into consideration. The estimation of the item parameters and the ability parameter is called test calibration (Baker, 2001). The values of the items parameters and ability parameters depend on the type of parameter model used. These include three, two and one parameter logistic models. These models provide mathematical equation for the relation of the probability of correct response to ability (Baker, 2001). However, each model employs one or more parameter(s) whose numerical value(s) define a particular Item Characteristics Curve (ICC).

Item Response Theory includes one, two and three parameters item response models. For dichotomous model, it can be unidemension or multidemension. The three parameter logistic model is however considered the most general and the others are subset of it. The 3PL model comprises difficulty of item parameter, item discrimination parameter and item pseudo-guessing parameter. When the pseudo-guessing parameter is removed from the three parameter logistic model, the two parameter logistic model is left and when the discrimination parameter is removed

from two parameter logistic models, one parameter logistic model remains. The one parameter logistic model has a particular and unique property; it embodies measurement when the term is used in a strict axiomatic sense (Algumalai and Curtis 2005). The model characterises a test item in terms of only one feature which is item difficulty. Using this model, an ability is estimated for each possible raw score between 0 and 100 percent score. All examinees who make the same raw score are considered to have the same ability regardless of which items they answered correctly to obtain that raw score (Ryan, Osborn Popp, and Rivera, 2002).

The 2PL uses item difficulty and item discrimination parameters. The model includes item-level information that reflects the information showing that some items discriminate more sharply between higher and lower ability examinees than others. The IRT model (1PL, 2PL and 3PL) can thus be illustrated as follows:

One parameter logistic model

$$P\ (\theta j/bi)\ = \frac{e^{(\theta j - bi)}}{1+e^{(\theta - b\ )}}$$

……equation 1.1

Two parameter logistic model

$$P\ (\theta j/ai, bi)\ = \frac{e^{a_i(\theta - bi)}}{1+e^{a_i(\theta - b_i)}}$$

……equation 1.2

Three parameter logistic model

$$P\ (\theta j/ai, bi, ci)\ = c_i + (1 - c_i)\ \frac{e^{a_i(\theta - bi)}}{1+e^{a_i(\theta - b_i)}}$$

……equation 1.3

Where P; (θ) is the likelihood that learner with ability θ answers a random item appropriately, a_i is the item discrimination, b_i is the item difficulty and c_i is the pseudo guessing parameter. The 2PL model is obtained when c = 0. The 1PL model is obtained if c = 0 and a = 1 or constant. In IRT, higher levels of information are produced when items have higher discrimination "a" parameters, and smaller lower asymptote "c" parameters (Harvey and Hammer, 1999). A "b" parameter defines how easy or how difficult an item is and an "a" parameter determines how

effectively the item can discriminate between highly proficient students and less proficient students. The guessing parameter "c" determines how likely the examinees are to obtain the correct answer by guessing (Enuwah and Akwa, 2014).

Item Response Theory places item and examinee's performance on same scale thereby facilitating standard setting and making it useful in criterion-reference score interpretation (Yen, 2002). Item characteristic curve is used in IRT to describe the relationship between the probability of a correct response to the item and examinees ability. The curve shows the hypothesised and reasonable relationship. Examinees with greater ability have a higher probability of answering the item correctly and those with lower ability are less likely to get the right answer. Before the 1980's IRT research focused mainly on the estimation of model parameters, the assessment of model-information fit, and the application of these model to a range of testing problems using dichotomously scored multiple choice items. Research on performance assessment, polytomous response formats and multi-dimensional traits began afterwards. All IRT models make assumptions about the way examinees will perform on test items (Demars, 2010). They also assume that to answer correctly, an examinees response will be based primarily on the examinees general ability with regard to the subject matter being tested and up to three characteristics or parameters of the item, depending on the IRT model used (Downing and Haladyn, 2006).

There are three basic assumptions that must be met before using IRT models in psychometric process. They are dimensionality, local independence and monotonicity. Dimensionality is the assumption of the number of trait or ability the items measured. The unidimensionality assumption means that only one trait or ability is measured by the items. Local independence assumption means that responses for different items are not related. An item does not provide any clue whatsoever in answering another item correctly. If local independence exists, a large correlation between two or more items can essentially affect the latent trait and thereby causing lack of validity. Unidimensionality and local independence are alike but not the same concept. When the assumption of unidimensionality is met, so is the assumption of local independence. However, the assumption of local independence can be met without unidimensional information as long as all aspects that affect the test results are taken into account. Figures 1.1 and 1.2 illustrate the relation between examinee with 0 ability (Ee) and responses to different items (il, i2, i3, i4) within two situations: dependent and independent.

The third assumption of IRT is monotonicity, it specifies that examinees with higher scores on the traits have higher expected probabilities for answering item correctly than examinees with lower scores on the traits. The relationship between ability and probability of a correct response can be depicted graphically and it is known as the Item Characteristic Curve. As shown in figure 1.3, the curve is S-shaped (Sigmoid/Ogive). Furthermore, the probability of endorsing a correct response monotonically increases as the ability of the respondent becomes higher. It is to be noted that theoretically, ability (θ) ranges from -∞ to +∞, however in applications, it usually ranges between -3 and + 3.

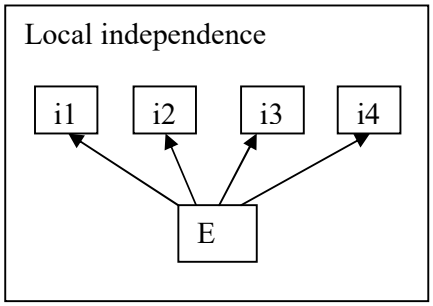

Figure 1.1: presents the illustration of independent items and examinee

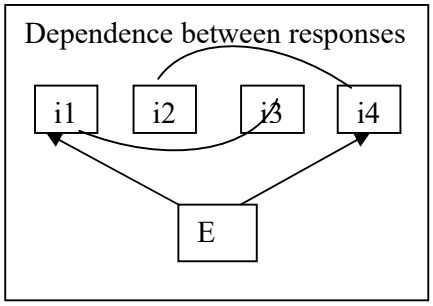

Figure 1.2: shows the illustration of dependent items and examinee

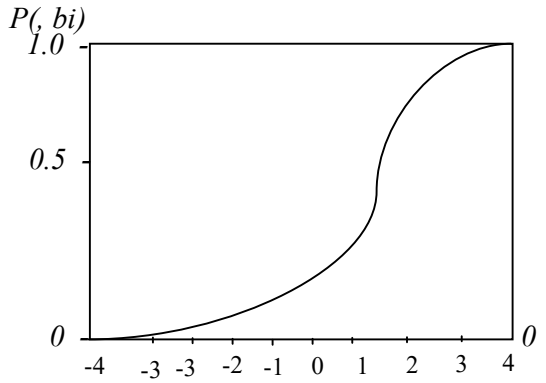

Figure 1.3: Example of an ICC from one parameter logistic model

In IRT test development procedure, identification of poor items is not straight forward as what obtains in the CTT. Generally, for a unidimensional model such as one, two, and three parameters logistic models, items which measure more than one trait of examinees ability are considered bad; items which are found to provide clue for answering another item correctly (that is item that are locally dependent) are also considered bad; items which misfit a chosen model that fit the test information are considered poor (Hambleton and Jones, 1993) and more importantly, items which are found to be biased in favour of or against examinees based on their group membership are also considered poor (Zumbo and Hubley, 2017). The issue of elimination of biased items will definitely confound the calibration of invariant item and person (ability estimate) parameters which are the hallmark of IRT framework.

However, biased test items refers to item that contains constructs that are irrelevant to the items. Biased test items result in systematically lower or higher scores for identifiable groups of examinees. Under both CTT and IRT frameworks, items on a test that are found to be biased in favour of a group of examinees are considered to exhibit Differential Item Functioning (DIF). Differential Item Functioning exists in tests when examinee of equal ability differs on average, according to their group membership, in their responses to a particular item. Group membership in this context refers to identifiable sub-populations such as gender, location of school, or other demographic variables to which test taker belongs. Prominent variables used for assessing DIF include gender (Ogbebor, 2016), ethnic group (Barnes and Wells, 2009), school location (Hogan-Bassey, 2011) socio-economic background (Nworgu and Odili, 2005). Moreover, four major approaches of describing and detecting DIF in a test that contains dichotomously scored items have been identified. These include logic regression analysis, Item Response Theory

method, log linear models and Mantel-Haenszel test (Welberg, 2007). However, only the Item Response Theory methods make use of item parameter estimates of DIF assessment while others are Classical Test Theory based (Oshima and Morris, 2008). In Nigeria, previous studies on DIF such as Ogbebor (2016) used the Classical Test Theory methods.

However, the best practices in test development require that the test developers some steps in considering the quality of test items. For example, for test items to be considered good, they must satisfy the conditions of unidimensionality and local independence. Also, the items must not exhibit DIF and should have moderate difficulty and discrimination index. Items with too low or even negative discrimination value are considered to be poor items. Furthermore, the effectiveness of each item is quantified by the amount of information provided by the individual items to the whole test. Theoretically, IRT overcomes the major weakness of CTT. Item response theory has the capacity of providing invariant items parameters and examinee's ability estimates. The invariance property of item statistics of IRT measurement framework to handle test equation (Hambleton, Swaminathan and Rogers, 1991) and Differential Item Functioning (Welberg, 2007) are better than CTT framework.

In addition, the invariance properties of the items and person statistics of IRT measurement framework makes it possible to solve important measurement problem encountered in testing (Ojerinde, 2013). Another most important concepts in the test standardisation are validity and reliability. The most important consideration in the use of assessment procedure is validity (Crooks, Kane and Cohen 1996). The proposed use of the test determines the degree to which evidence and theory support the test scores interpretation and this is evaluated instead the test itself. According to American Psychological Association (1996), test validation process involves providing a conceptual structure to a test through 'defining the skill, knowledge, intelligences, procedures, or features to be measured'. More importantly, is construct validity, the construct being measured must be clearly distinguished from other related constructs. However, the procedure of validation can lead to reconsideration of the test and the underpinning theoretical structure.

Test validation is the shared duty of the test designer and user. American Psychological Asssociation (1996) when the purpose of the test differs from what is supported by the test developers, the test consumer develops special responsibility for

test validation. Numerous kinds of evidence may be relied on to substantiate validity of a test. These include evidence based on:

i. test content, that is, analysis of connection among the test content and construct (domain) it planned to assess;
ii. the correlation of performance in a test to additional variables, such as particular criterion the test is projected to predict (predictive evidence) or not predict (discriminate evidence) and
iii. the outcomes of assessment, for instance, the effect of enlisting students in a learning experience (Jacob, Darrel and Paulgros, 2018).

American Psychological Asssociation (1996), reliability refers to consistency of assessment instrument on a repeated measure. The instrument is tested on a particular group of students over time. Close to this, is error of measurement. It shows 'the conjectural variance between the learner's real score on a specific measurement and the learner's score or general score for a procedure'. In giving evidence to aid test score reliability, users of test are required to discover the main causes of the error, summary statistics (descriptive statistics) to be employed with the magnitude of such errors, and the extent to which such scores can be generalised, test substitute, test administrations, scorers, and other pertinent information. Instruments for assessment of early reading literacy such as Test of Early Reading Ability (TERA), Early Literacy Skills Assessment (ELSA), Phonological Awareness Literacy Screening (PALS), Test of Phonological and Print Processing have been developed, standardised, produced and made available for public use in developed nations. However, the researcher is not aware of any Early Reading Literacy tests developed and standardised in Nigeria and made available for use in the pre-school centres in the country. The study, therefore, focused on the development and standardisation of Early Reading Literacy Test (ERLT) for preschool children in Nigeria.

In year 2012, at a writers' workshop on the development of training manual for Early Childhood Care and Development Education teachers in Nigeria organised by Universal Basic Education Commission, Abuja, as the participants got to the aspect of assessment of pre-school children, it was generally observed that there was no instrument that is developed and standardised in accordance to Nigeria culture and the approved early learning development standards for assessment of children at this level of education. The researcher was challenged by a Professor of Early Childhood Care and Development Education from University of Ibadan, that the Institute of

Education which specialises in educational and programme evaluation where the researcher was undergoing a postgraduate programme should be able to come up with standardised instrument for assessing pre-school programmes and children. The researcher, being a students of Communication and Language Arts and as gathered from literature that reading is the key to comprehension of other subjects and academic achievement, decided to carry out the study on the development and standardisation of Early Reading Literacy Test for pre-school children under the supervision of experts in the Institute.

1.2 Statement of the Problem

Early Reading Literacy Skills play a vital role in shaping a child and enhancing early learning experiences that are associated with academic achievement. Any child lagging behind in reading literacy skills development may likely find it difficult to perform as expected in his or her academic pursuit. Assessment of child's early reading literacy skills becomes essential aspect of holistic Early Childhood Care and Development and Pre-Primary Education programmes. However, this has not been given the needed attention in our society. This is revealed in observations from the series of monitoring exercises carried out by the quality assurance officers from the Universal Basic Education Board to some of the public primary schools where it has been found that a large percentage of pupils cannot read even a recommended English text in the class.

The problem probably is that, quite often, at the pre-school which is the foundation level, children were not tested in reading with standardised instruments based on the approved early learning development standards in language domain for Nigerian children. This might be due to non-availability of standardised test items that are capable of measuring pre-school children's latent reading ability. Consequently, the planning and decisions that are made about these children may be inappropriate, misleading and not uniform. This study, therefore, developed and standardised Early Reading Literacy Test (ERLT) for Nigerian pre-school children to assist those working with children in generating valid and reliable information on the reading ability of the children in order to take uniform and quality decisions.

1.3 Research Questions

The following research questions guided the study;

(1) Which of the IRT models for dichotomous data fits the ERLT?

(2) What are the estimated parameters of the Draft Early Reading Literacy Test item
(DRA-ERLT) using Item Response Theory framework?

(3) How many of the items of the DRA-ERLT survived using IRT framework?

(4) Do Draft Early Reading Literacy Test Items (DRA-ERLT) differentiate significantly between;
   (a) Gender (male and female pre-school children)
   (b) School location (Urban and Rural)
   (c) Type of school (private and public)?

(5) What are the items and person statistics in the DEV-ERLT using IRT framework?

(6) What are the estimates of reliability for sub-sets Early Reading Literacy Test?

(7) What are the range of difficulty levels of identification (of alphabets, objects, animals, birds, human beings and part of the body), recognition of signs and symbols and colours identification, reading fluency and picture reading subsets in the DEV-ERLT performance test?

(8) Are there normative data developed to facilitate the interpretation of the Early Reading Literacy (ERLT) scores with respect to gender, age, school type and location of school?

1.4 Scope of the Study

The study focused on the development and standardisation of Early Reading Literacy Test (ERLT) for pre-school children in Oyo State. It involved only the children at the final pre-school level in public and private schools in Oyo State. This group was selected because it is believed that they must have been exposed to some curriculum contents meant for children within that age and class in preparation for primary level of education.

1.5 Significance of the Study

Despite the fact that there are various reading assessment instruments, available literature indicated that very few of these instruments in Nigeria have their psychometric properties estimated, and virtually none is standardised for Nigerian children at pre-school level. This study, will therefore serve as the basis for further researches into standardisation of culturally appropriate Early Reading Literacy Test instrument for pre-school children in Nigeria. Instrument that was developed and standardised in this study will go a long way in solving the problem of lack of standardised instrument in assessing pre-school children's early reading literacy performance in Oyo State and generally in Nigeria.

This study is also valuable because the test developed and standardised will assist in having a uniform instrument of assessing and comparing pre-school children performances on early reading literacy achievement based on gender, school type, school location, age and even state. This will guide the stakeholders in adjusting and improving stimulation to learning package on early reading literacy skills for pre-school children for better performance.

1.6 Definition of Terms

1.6.1 Conceptual Definition of Terms

Reading Literacy: This refers to an individual capacity to engage in written text, understand, use and reflect on it in order to develop ones' knowledge and achieve ones' goals.

Item Parameters: These are estimates used to fit the model in this study. They include discriminating index and difficulty index.

Logit: This refers to the feature of IRT where test taker's characteristics (gender, type and location of school) and item characteristics (discriminating and difficulty indexes) are presented on the same scale in this study.

Item Response Theory (IRT): It is an attempt to model the relationship between an unobserved variable, usually conceptualised as an examinee's ability and the probability of the examinee responding correctly to any particular test item.

Latent Traits: These are the measures of innate, inherited cognitive characteristic of examinees that cause constant performance in a test and influence his or her response to the test items. These can neither be physically observed nor directly measured.

Fit: It is the measure of whether the items and the estimated latent traits are in

agreement.

Differential Item Functioning: It is a statistical characteristic of an item that shows the extent to which the item might be measuring different abilities for member of separate groups.

Test Standardisation: It is a process of producing valid and reliable test, administering it to a large defined number of test takers, norming their raw scores using percentile rank, T score and or other methods to generate norms (age norm, school location, school type and gender norms).

1.6.2 Operational Definition of Terms

Early Reading Literacy Test: This is an instrument developed by the researcher to assess the reading literacy skills of children at the final preschool level. It assessed five major sub-skills in emergent reading: (i) knowledge about print (ii) oral language ability (iii) phonological awareness (iv) picture reading and (v) reading fluency. For example, reading fluency is measured by scoring any response with word substitution, omission, wrong pronunciation '0'. For a test taker to score '1' the full sentence has to be read correctly.

Early Learning Development Standards: This is an acceptable range for measuring performances among pre-school aged children in Nigeria and for assessing their levels of school readiness. The expected performance of preschoolers across eleven selected domains is outlined. This study focused on reading literacy skills and the instrument was developed to assess pre-schoolers based on these standards. For example; A child in the final pre-school class should be able to tell story following picture which he or she sees for the first time; identify individual letters of the alphabet; recognises some signs and symbols in the environment; recognises own name and sings nursery rhymes and songs. If a child cannot perform any of these tasks, it indicates that such reading sub-skill has not been developed in such a child.

Final pre-school level children: According to this study are young children between ages four - six years that attend the final level of pre-primary class both in private and public schools in Oyo state.

Differential Item Functioning: It is the identification of the level of priority that should be given to the teaching of objects identification, reading fluency and picture reading of identifiable sub groups.

It is assessing the level of differential mastery of pre-school children in respect to

gender, school type and location that should be given more attention by the teachers while stimulating the children to learning. Items that showed DIF favour females calls for more attention to males by the pre-school teachers while stimulating them to learning and the other sub groups as applicable.

Item Parameters: According to this study, item difficulty index range is interpreted as follows:

Very Difficult Items: According to this study, does not mean they are bad items but it implies that pre-primary teachers should pay more attention and increase the stimulation of children to learning those aspects of early reading sub-skills.

Very Easy Items: According to this study very easy item indicates that pre-school children have really mastered those aspects of early reading skills and teachers need not to dwell more on them while stimulating the children to learning.

Item Discrimination: In this study, items with very low or low discrimination value and with very high difficulty value were rejected except items on the identification of chronologically arranged capital alphabet letters.

CHAPTER TWO

LITERATURE REVIEW

This chapter focuses on review of literature and theoretical framework related to the study.

2.1 Theoretical Background

2.1.2 Cognitive Theory of Literacy Development

Cognitive Theory of Literacy Development was propounded by Chall in 1993. The theory states that the art of being literate is basically taught and learned. For alphabetic languages, print is a code that represents phoneme/grapheme correspondence; therefore, learning to read and write begins with learning the code. In Cognitive Theory of Literacy Development, the stages of reading or writing development are necessary to guide teaching (Chall, 1993; Ehri, 2006). They affirmed the cognitive viewpoint as "independent" which means that literacy involves practical abilities which must be acquired individually from societal or cultural influences. Recently, phonological processing has been known as an essential component of reading literacy, and its development is also understood to occur in stages. An instance of suggested stages of reading development is given. Persons, even people with one disability or the other go through stages in acquisition of reading skills in certain age limits, and following the same order (Chall, 1993). Chall puts forward the six reading stages:

Stage 1: Birth to age six, it is called Preliminary-reading

Stage 2: Ages Six to Seven, known as Early Reading or Decoding

Stage 3: Ages Seven to Eight, refer to as Validation, Eloquence, Ungluing from Print

Stage 4: Ages Eight to Fourteen, known as Reading for Learning the New

Stage 5: Ages 14 to 18, known as Multiple Viewpoints

Stage 6: Ages 18 and above, refer to as Construction and Deconstruction.

Going over the six stages described, they have similar characteristics like the identification and decrypting of words, relating the verbal word to the printed materials, observing the rules concerning linking letters to sounds, learning the denotations of uncommon words (non-concrete words, ideas, concepts), and knowing word knowledge that is required for comprehending what is read. Chall's stages may be used to identify what an individual has learnt and what is yet to be taught. The theory also recommended norm-referenced tests to identify a reading problem. Gillon

(2017) stated that a number of studies adopting various methodologies and conducted in a variety of alphabetic languages has credibly established that a powerful connection occurs between phonological awareness and literacy development (Shaywitz, 2005). Phonological awareness acquisition also consists of a hierarchy of sub skills that progress from word level to syllable being a determinant of early reading performance, through phoneme and syllable. At the word level, readers are able to distinguish words within a sentence. Moving away from word discrimination is the capacity to know that words can be broken into smaller fragments such as syllables, onset and rime, and phonemes. Some theorists postulated that all the sub skills should be taught in order for reading ability to develop (Gillon, 2017), while others claim that phoneme acquisition is the most important factor for reading success (Shaywitz, 2005).

As regards to writing, since writing literacy cannot be totally separated from reading literacy development, Johnston and Johnston (2002) proposed that writing develops according to these steps:

(a) emergent (ages 1-7): drawing, scribbling, pretend writing, printing letter-like to actual letters; no sound-symbol correspondence;

(b) beginning (ages 5-9): early writing is laborious, but it improves to the point of accomplishing half a page of written work the content of which is often a summary or retelling;

(c) transitional (ages 6-12): more fluency, planning, organisation, and details characterize this stage; and

(d) intermediate and specialised writing (ages 10-100): fluent writing with expression and voice and varied styles and genre are seen.

Accompanying these writing stages are levels of spelling skills: pre-literate (emergent)- draw a picture or scribble and later write unrelated letters; early letter name (early beginning): writes predominant sounds in words and then early and final consonants; middle and late letter name (later beginning): use of early and final consonants with a vowel in most syllables, progressing to short vowel patterns, consonant blends and digraphs, some long vowel words; within-word pattern (transitional): spell short vowel words, most one-syllable long vowel words, r-controlled words, and use of some Latin suffixes; and syllable juncture and derivational constancy (intermediate); learn how syllables fit together, to double consonants, drop the *e* to add an ending, know suffixes and prefixes.

The Cognitive Theory of Literacy Development believed that learning to read starts with learning the code which is the alphabets and this study considered the identification of alphabets in English Language as the most important and first step in learning to read. The subtests in identification of alphabets had largest percentage of items in the developed and standardised instrument in this study. The theory highlighted six stages of reading literacy common to every individual at certain age, the first stage which is called pre-reading meant for children from birth to age six is the focus of this study. Cognitive Theory of Literacy also believed there is a strong relationship between phonological awareness and literacy development. The subtest on the pre-school child phonological awareness was part of the instrument developed and used in this study. The theory also recommended the use of normed referenced test to diagnose reading difficulties, this is relevant to this study because the reading literacy test developed was standardised across age, sex, school type and location norms.

2.2 Item Response Theory and Related Models

There were three major pioneers in the work of Item Response Theory. As cited in Hambleton and Rogers (1991), they are; Frederic M. Lord (an Educational Testing Services Psychometrician), George Rash (Danish Mathematician) and Paul Lazarfeld (An Australian Sociologist). However, the purpose of IRT is to provide a framework for evaluating how well the items in the assessment instrument relate to the individual test taker. It is mostly used in educational research, while psychometricians use it for developing and equating the difficulties of successive version of tests (Thissen and Orlando, 2001).

Ojerinde (2013) stated that IRT is also described based on the number of parameters in the model and there are four types of parameter model. They are:

(i) One-parameter model. This is also called the Rasch Model. This model assumes that all items discriminate equally among the testees. It is only interested in the difficulty level of the items.

(ii) Two-parameter model. This model considers the fact that items of a test cannot discriminate equally among all the testees. It estimates two parameters, “a” (discrimination index) and “b” (difficulty index) but assumed that a test taker cannot answer a question correctly by guessing.

(iii) Three-parameter model. This model assumes that a test taker can answer

an item correctly by guessing. Hence, in addition to estimating discrimination index "a" and difficult index "b", it estimates guessing index "c".

(iv) Four parameter logistics model. This is popularly called carelessness and the model assumes that there are some items so difficult that even with extreme level of a trait, not every examinee will respond to the item correctly. Moreover, increasing the number of parameter will improve the fit of a model, but at some point it may lead to complication (Reise and Walter, 2010; Wen-wei, Rong-guey, Yung-Chin and Hsu-Chen, 2012).

There are three basic assumptions that must be adhered to in using IRT model;

(i) Assumption about dimensional structure of the test information. The model assumes that one trait/ability is enough to explain the examinee's performance. So the model can be one dimensional or multidimensional depending on the number of traits the model is interested in.

(ii) Assumption about local independence. This means that the test taker's response to one item does not affect his response to another item. This assumption is met in one dimensional IRT model when the probability of a test taker's response pattern is equal to the product of probabilities associated with the test taker's score for each item.

(iii) Assumption of mathematical form of Item Characteristics Curve (ICC): The key issue in IRT framework is the relationship between test taker's latent ability and the probability that a test taker will respond correctly to certain item(s) (Duong, 2004).

Being the modern test theory, Item Response Theory (IRT) comes with some benefits over Classical Test Theory (CTT) theory which include the following: its approaches are different due to the fact that they are item oriented rather than test oriented, meaning that the approaches estimate statistics for each of the items in the measure or assessment as it relates to the capacity level of respondents instead for the items as constituents of the total test for a specific cluster of respondents. Also, the information produced from IRT offers itself to modern test formats in which each person responds to chosen item based on their ability levels and not to all items in the test. Measurement experts now adopt Item Response Theory measures over Classical Test Theory measures for developing and evaluating tests and assessment (Jacobson, 1997). This study adopted the Item Response Theory Framework, the data from the

developed instrument was subjected to one-parameter logistic model and two-parameter logistic model, the result revealed that the two-parameter logistic model had the lowest value which indicates that the two-parameter logistic model fits the Early Reading Literacy Test (ERLT) data obtained from the study, this made it possible to calibrate the test. Item discrimination index (a) and difficulty index 'b' were estimated for the 226 items in the DRAFT ERLT as well as the selected 163 DEV-ERLT. The dimensionality of the test data was also estimated using Stout's test of essential unidimensionality implemented in DIMTEST version 1.0 which showed the test is unidimensional, fulfilling one of the assumptions in using IRT model.

2.2.1 Item Statistics

Item statistics provide information about examinees' responses to each test item in order to help judge its effectiveness. Two characteristics of the items of most interest are difficulty and discrimination. Item analysis is a process that empowers test developer. Knowledge of item difficulties, item discrimination, and distractors can help a test developer make decisions about whether to retain items for future administration, revise them, or eliminate them from the test item pool (Levitov, 2015). Item analysis can also help a teacher to determine whether a particular portion of course content should be revisited. In any case, all indices should be considered together before making decisions or revisions.

Mean

Mean is the average of test scores to an item. It is computed by adding up the number of points earned by all examinees on the item, and dividing that total by the number of examinees.

$\frac{\varepsilon f x}{\varepsilon f}$.....equation 2.1

Standard Deviation

Standard Deviation (S.D) is a measure of the dispersion of examinee scores on an item. It indicates how 'spread out' the responses were. A large standard deviation means that there is much variability in the test scores of the group (i.e. test takers performed quite differently on the test). A small standard deviation means that there is little variability amongst the scores (i.e. examinees performed quite similarly on the

test). The item standard deviation is most meaningful when comparing items which have more than one correct alternative and when scale scoring is used. For this reason it is not typically used to evaluate classroom tests.

Item Difficulty

Item difficulty is also referred to as p value. For items with one correct alternative worth a single point, the item difficulty is simply the percentage of examinees who answer an item correctly. In this case, it is also equal to the item mean. To compute the item difficulty, the number of people who answered the item correctly will be divided by the total number of people who answered the item. For a large sample, add the number correct in the top group (Ru) to the number correct in the bottom group R_L) thereafter, divide the sum by the total number of examinees in the top and bottom groups (N). This is illustrated mathematically below:

$\frac{RU+RL}{N}$ ……equation 2.2

The item difficulty index ranges from 0 to 100 but more typically written as a proportion of 0.00 to 1.00, the higher the value, the easier the question. Item difficulty is relevant for determining whether students have learned the concept being tested. It also plays an important role in the ability of an item to discriminate between students who understand the tested material and those who do not. Table 2.1 illustrates the classification of item difficulty. Item difficulty is classified as 'easy' if the index is 91% or above, as a greater percentage of examinees got the item correct; 'moderate' if it is between 61% and 90%; 'difficult' if it is between 21% and 60%; and very 'difficult' if it is 20% or below, since only a small percentage of examinees got the item correct. Popular consensus suggests that the best approach is to opt for a mix of difficulties. That is, a few very difficult, some difficult, some moderately difficult, and a few easy. The item will have low discrimination if it is so difficult that almost everyone gets it wrong or so easy that almost everyone gets it right. Typically, items with moderate difficulty level (p=.5) are retained to increase score variability, ensure that scores will be normally distributed, provide maximum differentiation between examinees and to help maximize test's reliability (Thorndike, Hagen and Sattler, 2006). If the goal of testing is to choose a number of examinees, the preferred difficulty level will be equal to the proportion of examinees to be chosen. For

example, if only 15% are to be admitted, the average item difficulty level for the entire test should be 15.

Based on this study, after the analysis, table 2.1 shows the benchmark for Item Difficult Classification:

Table 2.1 Classification of Item Difficulty

% Correct	Item difficulty classification
0-20	Very difficult
21-60	Moderate
61-90	Easy
91-100	Very Easy

2.2.2 Differential Item Functioning

Differential Item Functioning (DIF) is a phenomenon that arises when the probability of answering an item correctly is independent of true ability but dependent on membership to a group. Technically, Differential Item Functioning occurs when an items between two groups functions differentially, - a reference group and a focal group (Nworgu, 2011). In the same vein, Zumbo and Hubley (2017), states that DIF occurs when different groups of examinees show differing possibilities of success on the item after matching on the underlying ability that the item is intended to measure. There are two kinds of Differential Item Functioning that can be displayed by an item. They are uniform Differential Item Functioning and Non-uniform Differential Item Functioning (Welberg, 2007). A situation where there is no interaction between ability level and group membership, that is, the probability of answering an item correctly is greater for one group uniformly over all matched ability levels is known as Uniform Differential Item Functioning. On the other hand, each item shows Non-Uniform DIF, if there is an interface between capacity level and group membership. For an item to display Non-Uniform DIF, the probability of answering an item correctly is not the same over all matched ability. Detection and estimation of interactions between item difficulties and various subgroups within the population of respondents is the main purpose of DIF. It is most often applied to interactions with respect to demographic or ethnic groups such as gender, location or race. (Welberg, 2007)

The relevance and irrelevance of constructs are some factors that can contribute to the interactions between item difficulties and various sub group within a population of respondents (Nworgu, 2011). When Differential Item Functioning occurs as a result of construct relevant factor; such as actual examinees ability differences, a real or true group differences called Item impact is suggested (Huff, 2000). This is an indication that the test is measuring what it is designed to measure, but when Differential Item Functioning occurs as a result of irrelevant factors such as demographic affiliations For example, socio-economic status (high on low socio-economic status), location of school (rural or urban) or gender male or female, a systematic error or item bias is suggested. Specifically, "item is said to be biased when test takers from one group are less likely to answer an item correctly than test takers of another group due to some characteristics of the item or the test situation that is not relevant to the purpose of the test" (Welberg, 2007). For example an item is said to be biased or exhibit Differential Item Functioning if the probability of boys responding to a specific test item differs from girls when they are both operating at the same overall level of ability on the measured construct. Also, item impact according to Dorans and Hollands (1993) refers to situations when test takers from different groups have different probabilities of responding correctly to an item due to true differences in ability measured by the item.

It is important to note that when assessing the Differential Item Functioning test items, two groups of test takers are usually involved (the reference group and the focal group). The focal group is the group of interest of examinees which is usually the minority group. And the reference group is the group to which the focal group performance on an item is compared (Angoff, 1993; Welberg, 2007). It is possible to test DIF among more than two groups simultaneously (Barnes and Wells, 2009). In literature, several methods of describing and detecting Differential Item Functioning have been advanced. The choice of which DIF methods to be applied for assessing the DIF of items on a test depends on the type of items contained in the test. (Zumbo and Hubley, 2017; Welberg, 2007), When a test contains dichotomously scored items, four methods of DIF are applicable. They are: Logistic Regression (LR), IRT methods, Log Linear models (LLM) and the Mantel-Haenszel (MH) test. These DIF analyses except the IRT methods, do not utilize Item Parameters estimates because they are CTT-based (Oshima and Morris, 2008).

Item Response Theory (IRT) has a number of methods for identifying DIF because there are several IRT models. In Item Response Theory, DIF is assessed by comparing the Item Characteristics curves (ICC's) of different groups on an item. If the ICCs for each group are identical or very identical, it can be said that the item does not display DIF. If, however the ICCs are significantly different from one another across groups, the item is said to show DIF (Zumbo and Hubley, 2017). IRT procedure can be used to detect both uniform and non-uniform DIF. Using IRT, an item reveals uniform DIF when the ICCs for two groupings are dissimilar but parallel, while an item shows non-uniform DIF when the ICCs for two groups are different but not parallel (Welberg, 2007). The area between the two groups ICCs gives a hint of the degree of DIF in the items (Camilli, 2006; Camilli and Shepard, 1994; Swaminathan and Roggers, 1990). An item that shows DIF is considered biased. However, DIF alone does not render an item invalid (Nunnally and Bernstein, 1994). Item that show Differential Item Functioning performance between two groups of examinee are content appropriate and valid. Therefore, before an item is removed from a test based on DIF, the item has to be subjected to subject matter expert to know what actually went wrong with the item for judgment of removal if it unnecessary when only a few items of the test exhibit DIF, they are usually removed without impairing measurement of the intended construct (Wiberg, 2004).

Information needed for describing and detection of Differential Item Functioning in the process of test development consists of identifiable sub-populations that almost all the test takers belong. As noted earlier such sub-populations include: Gender either male or female); demographic variables, such as school location (either rural or urban). Relevant information on group membership can be collected for DIF detection by administering background questionnaire along with tests. This study subjected the data obtained to the detection of DIF based on the examinees gender, age, type of school attended and its location as gathered from the Section A of the test instrument on the demographic data of each examinee. Few items exhibited DIF in favour of males and females, some in favour of examinees from urban and rural. The researcher did not reject any of these items since they were found to be contextual appropriate. The implication of this study is that the pre-school teachers should continue to stimulate the children to learning on any item that shows DIF until all the children irrespective of their subgroups are able to perform such a task as expected.

2.3 Conceptual Review

2.3.1 Early Childhood Care and Development Education Curriculum in Nigeria

The Federal Republic of Nigeria (2014) promotes the wellbeing of children through the provision of education contained in the National Policy on Education (NPE). The Policy stated among other things that the purpose of Early Childhood/Pre-Primary Education is to effect a smooth transition from the home to the school, prepare the child for the primary level of education and teach the rudiments of numbers, letters, colours, shapes, forms and so on through play. A child cannot be taught outside the context and demands of his society. In order to fully incorporate the child into satisfying the societal demands, he or she is basically taught through various forms of play activities. It is as a result of this that the NPE lays emphasis on the use of culturally appropriate curriculum in educating the child at the early childhood stage and beyond. Early Childhood Education and Pre-Primary education is meant to cater for the diverse and numerous needs of the child, ranging from the physical, intellectual, socio-emotional to the aesthetic. In order to achieve this, the Nigerian Educational Research and Development Council (NERDC) came up with a curriculum guideline to meet these needs.

In trying to actualise these stated objectives, play method is considered most appropriate. For these purposes, the curricula of Numeracy, Literacy, Social skills, Introduction to Science and Technology, Engineering and Mathematics (STEM) make sense to the children when they are taught in a way that would relate learning to real life experience. In some cases, as rich and robust as the content of most of the subjects are, the child may not have the opportunity of applying all that he has been taught to his immediate life conditions. This may be due to the fact that either enough time has not been spent in the process of teaching or that the method of teaching may not be appropriate. There is this consciousness on the part of the teachers that they have limited time to spend on a particular topic and in view of this, they have to meet up with the timing of the scheme of work whether the pupils comprehend it or not.

Another constraint identified is the issue of compartmentalisation of subjects is that as much as the subjects are relevant to the holistic development of the child, their compartmentalisation contradicts the play method as recommended by the policy. For example, in the acquisition of numerical knowledge through play, it may not really cross the mind of the children or the teacher that other skills can also be

incorporated other than teaching those numbers. For example in a rhyme:

One-two buckle my shoe (knowledge of responsibility and personal care – social skills)

Three-four knock at the door (act of courtesy – social skills)

Five-six pick up sticks (safety/caution – social skills)

Seven-eight lay them straight (aesthetic value – STEM skills)

Nine-ten a big fat hen (scientific skill of observation/measurement – STEM skills). The child has learnt how to count, be courteous, be responsible, sensitive to sounds in language, the scientific skills of observation and measurement. As remarked earlier, the separation of subjects limits the possible uses of resources. Opportunities may not be available to the learners to draw out other subjects and skills from that singular class activity. This has called for the use of Integrated Early Childhood Development Education Curriculum as compiled by National Educational Research and Development Council. It ensures that learning takes place in a relaxed frame of mind before the test/test consciousness gets incorporated (National Educational Research and Development Council, 2005).

2.3.2 An Integrated Early Childhood Education Curriculum for Nigeria

An Integrated Curriculum in Early Childhood Education is one which is relevant and meaningful to the learners. It addresses all areas of child's development-physical, cognitive, emotional and social domain. An integrated curriculum is developed to rectify observed shortfalls in the former curriculum which is replete with subject barriers and compartmentalisation or pigeon-holing of knowledge. An integrated curriculum promotes continuity in learning across different subjects to promote holistic development. Shoemaker (1989) defined an Integrated Curriculum as the one that is organized in such a way that it cuts across subject-matter lines, bringing together various aspects of the curriculum into meaningful association to focus upon broad areas of study. It views learning and teaching in a holistic way and reflects the real world, which is interactive. Drawing an inference from this definition, an Integrated Curriculum should be organized in such a way that subjects merge into one another; that is, subjects are not taught in isolation. This is to prevent boredom and exploit short children's attention span optimally. Moreover, acquisition

of necessary skills for a meaningful life is possible for only those who are able to put knowledge to work in real life situations.

Integrated Curriculum is a blend of content areas into thematic or problem-focused units of study and a child-centred approach to learning and instruction. Since learning is particularly meant for the holistic development of the child, the curriculum should also be holistic in nature, unifying all the subjects in order to promote comprehensive learning. This is contrary to what applies in contemporary schools where teaching and learning is predominantly done by memorization, recitation of facts and figures. The word integration implies elimination of separation, rigidity, and inflexibility. Akinpelu (1987) stated that integration implies flexibility, unity and gives room for adjustments and re-adjustments until the goals outlined are achieved. In other words, subjects are not taught as independent whole but avenues are created whereby subjects are unified to help produce continuity in learning. In actual fact, both children and adults do not learn real life events in compartments but as a whole, there is always room for linkage. An Integrated Curriculum helps the child takes charge of his own learning, promotes creativity, brings novelty into teaching and facilitates learning. Adopting an Integrated Curriculum creates room for engaging in a combination of subjects, emphasising project, flexible schedules, flexible children grouping and its teaching and learning goes beyond textbooks. It also establishes continuity in what children learn in different subject areas. Our daily lives incorporate the use of Arithmetic, English, Science, Music and Rhymes to tackle life's challenges. The real world does not work in isolation neither does it work with individual subjects. It thrives on the holistic nature of knowledge. The use of curriculum integration increases the relevance of learning experiences because children are able to relate the knowledge acquired in diverse subjects (Akinpelu, 1987).

2.3.3 Components of Early Literacy Curriculum

Many educators and policy makers agree that a strong start in early literacy is critical to learning achievement but little agreement concerning how this is best achieved. A main worry is making sure the syllabus focuses the general growth including children's learning through emphasising the physical, social, emotional and general intellectual growth of children as well as strengthening the academic achievement. The stakeholders opposed such curriculum that dwelt too scantily on

literacy skills and disregard attention for all the areas of development that have things to do with promoting children's personal and academic growth. Certainly, the physical, social, emotional, cognitive and language development of little children are the major aspects that impact early literacy growth (Strickland and Schickedanz, 2004). A basic curriculum will serve the children differently. The children differ in what they bring to the pre-school situation and what they benefit from it. Many children come to pre-school having got wealth of experiences from storybooks and other printed educational resources, visited exciting places, involving in problem-solving activities, engaging in stimulating discussions and events that intended to increase their knowledge web and academic growth. The acquired linguistic and previous knowledge will make such children to gain from a curriculum that strengthens and increases the rich reservoir of skills and knowledge these learner possess. Other children require additional, different or specifically targeted learning opportunities in pre-school. Resourceful teachers and the specialists who advise them, make adjustments within the framework of the curriculum to make instruction more responsive to children's needs (Tabor, Snow and Dickinson, 2001).

Roskos and Vukelich (2006) stated that if a curriculum is home-produced or academically produced, those who design and use it are required to back up their claims with a research base. However, the main constituents of the then literacy curricular written on evidence-based early literacy studies comprise: (1) spoken/oral language development, this includes vocabulary and listening; (2) understanding of the alphabetic code, it includes phonological/phonemic awareness and knowledge of the alphabet; and (3) knowledge and understanding about printed materials and its use.

Spoken/Verbal Language: Verbal language grows simultaneously with literacy development, which comprises listening, vocabulary development and oral expression. Verbal language development is aided when child has various chances to use language in conversation with the elderly ones with other children coupled with listening and responding to narratives. Little children build vocabulary web when a they involve in events that are cognitively and linguistically inspiring by motivating them to define events and build prior knowledge.

Alphabetic Code: English language is an alphabetic language, which means that the letters we combine to write symbolise the sounds of the language that we speak.

Awareness of the alphabet letters and phonological knowledge (the ability to differentiate the sounds within words) form the basis of early decoding and spelling ability, both are connected with later reading and spelling success. Young children can learn to name letters and to differentiate them from one another. They can also begin to develop knowledge of the different sounds within words, such as syllables, rhymes and phonemes.

Children should be immersed in language-rich environments in order to develop phonological awareness and similarly, it would be difficult to master the ABCDs without lots of exposure to the alphabet (in books, on blocks, on refrigerator magnets, in cereal, in soup and so on). ABCDs awareness with speech sound knowledge do not frequently just occur from exposure for most children, though, parents and older brothers and sisters often deliberately teach children the alphabet, and study has shown that it is likely to teach phonological consciousness to youngsters and kindergarten children in ways that do not interfere with a complete and rich curriculum focus but do enhance later literacy (Kaderavek and Pertimonten, 2014).

Prints Knowledge and Use: Deducing meaning from printed materials includes awareness, knowledge and comprehension of the meaning of prints, such as the point to start to read a book or a page. Each of these is likely acquired from conversing with others around print. An early literacy curriculum might contain grocery store visits; reading to one everyday; having a lettering class where young children can do hands-on with printed message, and environmental print that is focused such as symbols, labels and diagrams. Furthermore, efficient early literacy teachers model the reading and writing processes during cooperative writing and reading. They clearly comment aloud concerning what they think when they read and write in order to make the process clear to children (Roskos and Vukelich, 2006).

2.3.4 Assessment Strategies in Early Childhood Care and Development Education

Assessment is one of the most important aspects of teaching. Teachers must repeatedly gather, synthesise and explain information concerning the learner. They should identify the true level of skill and knowledge of learner before teaching plan starts, again, evidence which shows learners are learning or not is required. This evidence is mostly based on the learner's observation and watching of pupils in the

class (such as. the richness of pupils' written work, their answers to questions). The information may be employed to achieve range of benefits:

(i) to guide and design future teaching;

(ii) to adjustor students' styles of learning, skills, interest, and students motivations

(iii) to give feedback and inducements;

(iv) to place students in learning groups and

(v) to identify challenges that pupils may be facing (Airassian, 2001).

However, the most common assessment tools are tests. Tests have a variety of types, such as puzzles, game, jigsaw, it can be printed, verbal or hands-on. In testing, sampling of some examinees' knowledge or intelligence is very important because it is on the basis of this an interpretation is made concerning his/her likely performance. The conclusion, in turn, may be adopted to take judgment about an individual or group of examinees (Madaus, Rusell and Higgins, 2009; Osterlinde, 1989; Crocker and Algina, 1986). This study also examined some forms of test in Early Childhood Education.

Standardised Tests

Standardised assessment is constructed to assess learner's attributes. The instrument may be given to an individual or a group of examinees. The purpose of standardised tests is to measure abilities, performances, intelligences, interest, attitudes, values, and personality characteristics and norm the scores in groups. Test outcomes may be used to design teaching, to examine differences among students, in addition, it is used for corrective measures (Wortham, 2008). There are many types of standardised tests for use with young children. Intelligence tests, performance assessments are considered as ability assessments because they assess aspects of capacity. Capacity refers the present level of knowledge or skill in a particular area (Wortham, 2008). Little children are frequently assessed in order for the purpose of gauging their learning progress. A test used for such children may target motor and language abilities, social or cognitive skills. For example, McCathy's Scales on Learner's Abilities (McCarthy, 1983) has indexes for spoken, perception – performance, numerical, reasoning and motor skills.

Intelligence test is an ability test and also an intelligence test due to the fact that it assesses ability for learning and problem-solving. The Stanford-Binet

Intelligence Scale (Thorndike, Hagen and Sattler, 2006) is an example of an intelligence scale that measures individual intelligence. However, success is related to the degree a person has acquired certain information or has mastered identified skills (Worthan, 2008). The Pebody Performance Assessment- Reviewed (American Guidance Service, 1997) is a measure of performance in Mathematics, Reading Recognition, Comprehension, Spelling and General Information. Intelligence tests measure the outcomes of overall and unintended learning and predict future learning. Specifically, ability refers to capacity to acquire knowledge and to increase expertise in other areas if such training exists. Like performance test, intelligence tests also assess learners' abilities (Wortham, 2008). Other types of standardised tests include the following;

Tests Meant For Screening

This is a test conducted to identify symptoms that a learner might have a reading difficulty which requires additional test. Screening tests is likened to diagnostic test which studies probable problems in order to know the steps required to fix the difficulties identified. The Denver II, for instance, may be used with babies and older children. Contrastingly, questionnaire on ages and stages was used for parental reporting (Squires, Brickler and Twombly, 2012). The parent can complete the questionnaire or participate in an interview with an examiner. It is administered from age four months to 60 months. The growth pointers in learning assessment (DIAL III) (Mardell - Czundo WSKI and Goldenberg, 1998) is also used for overall growth deferment. It is administered to the children between the ages three to six. The assessment contains researcher's personal observation with tasks given to the children. The Early Screening Inventory – Revised (ESI-R) (Meisels, Marsden, Wiske and Henderson, 1999) had two types: one meant for ages three to fout and a half, and the other was for a ges four and a half - sixyears. It screens developed cognitive domains and uses cut-off scores to determine whether the child requires additional assessment. Questionnaire for parent was used to gather additional information. First step screening Test for Evaluating Preschoolers (Miller, 1993) is another screening test. It has 12 subsets grouped into cognitive, communication and motor categories. There is also an Optional Social – emotional scale and adaptive behaviour check list which is conducted for children from ages two years, nine months to six years, two months.

Diagnostic Tests

Test for diagnostic assessment is administered once the learner has been screened and discovered there are signs that additional assessments are required. Measures of adaptive behaviour assess possible cognitive problems related to learning disabilities. Adaptive behaviour instruments attempt to measure how well the young child has mastered everyday living tasks such as toileting and feeding. The Vineland Adaptive Behaviour Scale (Sparrow, Balla and Ciechetti, 1984) assesses the everyday behaviours of the child that indicate level of development. The scale determines areas of weakness and strength in communication, daily living, socialization and motor skills. Another instrument in Adaptive Behaviour Scale-School (ABS – 5:2) (Nihira, Lelands and Lambert, 2014) assesses adaptive behaviour in 16 domains for social competence and independence. However, kindergarten intelligence test and adaptive behaviour scales are used to diagnose cognitive retardation. Although intelligence measures during the preschool years are generally unreliable because children's IQS can change enormously between early childhood and adolescence, they are used with young children to measure learning potential (Worthan, 2008). Other instruments such as the Kaufman Measurement Battery for Children (K - ABCD) (Kaufman and Kaufman, 1983), Batterle Develop Cognitive Inventory (BDI) and Bracken Basic Concept Scale – Revised (BBCS - R) address all domains of growth.

Tests of Language

Type of language tests for kindergarten remains essential due to the fact that many of them who are at risk of not learning effectively because they are poor at Language or their first language is not English are often examined prior to kindergarten. The Pre-school Language Measure (Zimmermann, Steiner and Pond, 2002) and Peabody Picture Vocabulary Test (Eigsti, 2013) provide information on a child's language ability, which can help determine whether a child will benefit from a language enrichment programme. The Preliminary-language Assessment Survey (Pre-LAs) developed in 2000 assesses verbal language skill. It is also used to make decisions on students' placement, scale improvement over a period of times and ascertain learner's wants.

Norm-referenced and criterion –referenced tests

According to Wilson (1980) norm-referenced and criterion –referenced tests are both standardised instruments. Certain standardised instruments are constructed targeting norm- referenced purposes and others for criterion – referenced purposes. The present development is to construct item for two types of test. The two test types have diverse purposes and test items are used differently depending on what one is looking for when gauging students' performance. Norm-referenced tests provide information on the performance of an individual examinee compare with that of others in a sampled group. Each learner's performance was compared with the group. The person's percentile rank is to determine the relative standing in a norm group by recording what percentage of the group obtained the same score or a lower score. In contrast, criterion – referenced test provides information on how the individual performed on some standards or objectives. The test outcomes permit users to interpret what an individual examinee can achieve without considering the performance of others. Criterion – referenced tests are designed to measure the outcomes of instruction, they determine the individual's performance on specific behavioural or instructional objectives (Wilson, 1980).

Linn and Muller (2005) described the difference between the two types of tests as the ends of a continuum. The criterion- referenced test emphasizes description of performance and the norm-referenced test emphasizes discrimination among individuals. Regardless of whether tests are norm or criterion referenced, the process of their design, development and standardisation follow all procedures of reliability and validity. There is even the possibility of norm-and criterion referenced test have not been standardised, however, criterion – referenced tests are more often non-standardised (Goodwin and Goodwin, 1993). It is equally important that criterion – referenced tests have validity and reliability, if they are to be used to make decisions about young children.

Criterion-and norm- referenced tests share similar features according to Linn and Muller (2005) described them in these ways:

(1) they both need a pertinent and representative sample of test items.
(2) they both need requirement of performance domain to be measured.
(3) they both employ the same kind of test items.
(4) they both employ the same rules for item writing (except for item difficulty)
(5) both are judged by the same qualities of goodness (validity and reliability)

(6) both are useful in educational assessment

However, the objectives for measurement are different in both tests. The norm-referenced test is broad in content and because it is concerned with overall performance, only a small sample of behaviours for each objective can be assessed. The criterion referenced test on the other hand focuses on mastery of objectives. Each objective has many test questions to determine whether the objective has been mastered. Performance test in Mathematics provides a good example. The norm referenced test for the elementary school may have items on addition, subtraction, sets and all other areas included in the Arithmetic curriculum. Test items are writing to sample the learner's overall performance in Arithmetic. The learner's total raw score is then transformed to compare overall performance with the test norms. While in criterion – reference test, learner performance on individual curriculum objectives is crucial. Test items are designed in assessing if the learner has mastered a particular learning objective in subtraction, addition, or other aspects of the Arithmetic curriculum (Goodwin and Goodwin, 1993). Furthermore, another difference between norm and criterion relates to differences in test items. In norms-referenced instrument, test items must cover a wide range of difficulty, because the test is intended to discriminate between the performance of students and groups of students, difficulty of test items ranges above the score level for which the test is intended. Test items designed primarily for criterion – referenced purposes are written specifically for learning tasks. Easy items are not omitted, and purpose is to evaluate how well the learner has learned the objectives for the score level (Wilson, 1980).

Norm – referenced test scores are used to measure individual performance within a designated group. Norms are not standards to be reached; they are numerical descriptions of the test performance of a group of students. Norms can be established at a national or local level. Norm referenced tests are commonly used to measure school performance, intelligence and personality traits. Formative assessments are administered at the pre-school level to identify children who need or can benefit from special instruction, as well as to determine the success of an early childhood programme (Wortham, 2008). Intelligence measures like Wechsler Kindergarten and Primary Scale on Intelligence (Wechsler, 2002) are norm – referenced tests that allow test examiner to differentiate the knowledge/skills of preschool children. Other tests for this set of children include the Kaufman Assessment Battery for children (K-ABC) (Kaufman and Kaufman, 1983) Scales of children and McCarthy's Abilities

(McCarthy, 1983). Intelligence tests are used to identify the gifted children in addition to identifying children with disabilities.

Criterion – referenced test scores are used to describe individual performance on specific objective. Criterion –reference do not measure the differences between individual performances, rather, they show if the individual examinees already have the mastery of the objectives tested. Criterion referenced instrument/test is used for screening, diagnostic assessment and instructional planning. In the preschool years, diagnostic assessments are the criterion referenced test used most frequently. Although diagnostic screening is used primarily to identify children who might profit from early education intervention or from special services before kindergarten or first elementary school level. It is also used as a checkpoint for children who are developing normally (Wortham, 2008). The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) (Good and Kasminski, 2002), is an example of a screening for literacy skills. One of the subtests of the DIBELS is Individual Sound Fluency (ISF). It is in a measure of phonological awareness and it is administered to preschool and kindergarten children.

More importantly, when teachers and other professionals conduct assessments with preschool children, they need to be sensitive to the special requirement of working with very young children. Young children have very short attention spans and are easily distracted. Administrators of assessment instruments and other strategies can benefit from the following guidelines;

i. get all resources ready before the assessment period and appraise processes for administering the assessment before the child comes;
ii. ensure the examinee is acquainted with the test environment when administering the test. The session might need to be conducted in the home for all the children. Test conductor must be familiar with the children also;
iii. create rapport with the children prior to the assessment period. Have a conversational exchange with the learner before the session begins, an interesting topic may do the magic. By the time the learner appears relaxed, the number one assessment/ tasks can commence;
iv. be sensitive to signs of fatigue or behaviours that indicate that the child is no longer responding to the assessment tasks. Take a brief break or remind the child how to respond to tasks before resuming the session;
v. use assessment time efficiently. The child should not be hurried, but

assessment tasks should be administered with little lag in time while the child is alert and attentive and

vi. study modifications that might be required for the examinee with incapacities. Be well-informed regarding how task might be modified within requirements for how standardised tests should be conducted. If alternative procedures can be used, allow the examinee to respond differently to a test item. However, caution must be made not to change the intent of the item or the type of response that is appropriate as well as correct (Wortham, 2008).

Whether formal or informal assessment strategies are employed, early learning standards require teachers and other professionals to plan ahead the assessment of learners. Standards need to be integrated into the existing curriculum and assessments that are proven to be of high quality for preschool children. Otherwise, they might find themselves narrowing the curriculum using inappropriate testing methods (Groulund, 2006; Oliver and Klugman, 2006).

2.3.5 Assessment process in Early Childhood Care and Development Education

In Early Childhood Education, assessment occurs throughout the school year. This section describes how assessment of the learner will commence early right from the start of an academic session to the final (summative) assessment (Wortham, 2008).

Beginning of the Year Assessment

Each year, according to Wortham (2008), once a teacher gets a raw set of children, the number one assignment is to observe their individual peculiarities and identify each child's level of growth. Kindergartens have irregular rate of growth. The physical, social, cognitive, and language develop differently for the children. Growth happens in spurts and may delay for a long period of time. The teacher can employ direct observation, checklists, and discussions with the child and parents to identify each child's present level of such domains. This first (formative) assessment affords the teacher with a starting point of development instruction and activities.

Ongoing assessment

According to Bredly-Johnson (2001), on-going assessment is conducted continuously throughout the year. While different activities are going on in the centre, the teacher notes the child's progress or difficulties that might hinder progress. This

information is made in anecdotal records or some other types of record-keeping system that can be used for planning. Assessment can be formative or summative. The only strategy the teacher uses to monitor a learner's progress in mastery of information or skills during a series of learning activities is formative assessment. Specifically, this may be given intermittently, while teaching specific ideas one after the other in order for the teacher to know if the particular objectives have been understood and which requires additional work or activities. (Breadly-Johnson, 2001)

Summative Assessment

The comprehensive assessment and progress report take place after the close of the school calendar. In this stage, the teacher summarises the learner's performance during the academic session. A variety of approach might be employed to determine students' progress, including students- designed assessment in different content areas, standardised performance tests, learner self-assessment, home-work and a written narrative of the students' accomplishment. The report is however handed over on to the subsequent teacher who uses the information during the commencement of another academic session (Wortham, 2008).

The test that is conducted at the end of learning activities is called summative assessment to confirm mastery of what was taught. Summative test conducted at the concluding end, the purpose is to determine whether the intended learning outcome has been achieved by the students based on learning objectives of the particular topic treated.

Assessment at the end of reporting periods

Generally, at the end of reporting period, teachers are asked to evaluate a child's progress and accomplishments. At this time, the teacher might record the child's progress for the period of time, as well as the plans for the child in the next reporting period. Feedback in verbal or written form is given to parents at the end of the term after the reports have been given to their children, the documented children work samples are at times included in the summary of report. (Oduolowu, 2011) In addition to observing the child, class teacher may adopt one on one discussion with learner in determining the child's view of the information/topic treated in the class. Furthermore, the child might have the opportunity of self- assessment while parents can also describe the observations of the child's progress.

2.3.6 Procedure for Development of Standardised Performance Test

Performance test is to find out the degree at which a learner has attained certain information or has grasped certain skills related to specific prior instruction. Performance tests are designed to measure the academic progress students have made over a period of time, it assists in identifying the appropriate academic placement for a learner and determining a young child's readiness for primary education. When constructing a test, the designer should follow series of steps, so as to ensure that the test achieves its purpose(s);

First, according to American Psychological Association (1996), the purpose of the test should be established. The purpose of every standardised test should be clearly defined. A clearly specified purpose of the test represents the framework for the test construction. A test description outlines the purpose of the test and the targeted examinees. Test developers need to consider how the test scores will be used whether for certification purpose, licensing the testees, testing of minimum competency or mastery of a specific subject. A standardised test whose purpose is clearly defined, can be easily evaluated at the completion of its construction. The test constructed in this study is aimed at assessing the early reading literacy skills of children at pre-schools. American Psychological Association (1996) stated the guidelines for including the test's purpose in the test manual. The test manual should state explicitly the purpose and applications for which the test is recommended. A standardised test may intend to measure cognitive, language and social development in the test takers, another may aimed at measuring oral language proficiency and assessing learner's needs while another may target identifying children for placement in primary education programme and other may intend to assess children's success in playgroup abilities and efficacy of head start programme.

A standardised performance test result can be used to evaluate test taker's performance in a particular subject, to determine each test taker's progress and to compare performance. The test results can also provide analytical information about learner's need for upcoming instruction and also to determine the effectiveness of instructional programme. A standardised assessment ought to define clearly the psychological, educational and other reasoning underlying the test. The construct the test will measure; to whom the test will be administered (population) should be clearly stated. A standardised test design for children will be different from the one design for adults. If the test is designed for children, their composition and

characteristics should be clearly defined. In test development, variables like age, educational level, socio-economic and cultural background should be included. The test constructed in this study is for pre-school children at the final level, characterised with attending pre-schools either in the public or private settings. Definition of the construct or domain (capacity, skills and body of knowledge) that is to be measured is very crucial. In the case of test that centres on performance, it entails the analysis of curriculum, instructional resources including text/workbooks. After the analysis, the domains to be measured will be presented in a table of specifications describing the content grid, that is, the expected learning outcomes of the subject matter (Bloom, Hastings and Madaus, 1971).

More specifically, a test blue print is a specification of what the test should cover rather than a description of what the curriculum cover. A test blueprint should include the test title, the purpose of the test, the aspects of the curriculum covered by the test, an indication of the examinees for whom the test will be used, the types of task that will be used in the test and how these tasks will fit in with other relevant evidence to be collected. It also covers the condition through which the test will be given (time, place, who will administer the test, who will score the responses, how accuracy of scoring will be checked, and any precautions to ensure that the responses are only the work of the test taker attempting the test), and the balance of the questions. Although, an assessment instrument may cover little section of the objectives that students' mastery is required, nevertheless it is very essential that the tasks/questions chosen for the test give sufficient depiction of the curriculum. Third, there is a need to determine the test format. Brown (1983) described this as the question format the test designer is going to adopt to present the test items as well as the response format, either verbally or in written form. Adults are familiar with written tests while young children can be tested orally by a test administrator. The response may also be true-false, objectives or matching items and short answer with the guide of a test administrator.

Test developer need to develop experimental test forms to be guided by test purpose description in delimiting the test content. The items of the test instrument must measure the purpose of the test and as well be accordance with the curriculum. For example, while developing performance test for children, the curriculum should be analysed, this is because the test should reflect the instructional programme. While constructing performance test for children to be used nationally, textbook series,

syllabi and curricular materials will be studied in ensuring test objectives reflect curriculum trends. Experts in the field should be consulted for validation. Burrill (1980) suggested that pool of items than will be used should be written, because many will be eliminated in the process of editing and re-writing. The process of developing good test items involves writing, editing, trying-out, re-writing or revising test item. The selected experimental test items would be assembled for trying out with a sample of children. This test resembles the final form of the instrument and should include instruction for administration. This is described as item tryout and analysis. Item analysis involves studying the characteristics of each test item, that is, the difficulty level, discrimination and score progression of difficulty.

The difficulty level of a test refers to how appropriate is the test in the sense of its difficulty with respect to the test takers. It is to find out the number of examinees in the pilot group that answered a particular item correctly. Test taker's level, age and purpose of the test are considered before fixing the difficulty level of a performance test. The difficulty level of a standardised performance test should not be too high or low. Test discrimination is the degree to which the test distinguishes the test takers into various groups according to their abilities. It involves comparing a person’s score with the average score of some relevant group of people. The score progression of difficulty refers to tests that are taken by students in different schools. Once a test item has good score progression of difficulty, reasonable proportion of students should answer it correctly in each successively higher score (Burill, 1980). Assembling of the final form of the test after the items have been re-examined, rewritten or eliminated is the next action in test standardisation. Test developer ensures that items selected actually measures the required behaviour. After assembling the test, instructions for both test takers and administrators, like, information about the testing environment and testing providers are written and printed along with the test instrument.

The final test form must be administered to another larger sample of test takers so as to generate norm-data. Norms give room for the comparison of children's test performance with that of a reference group. The norming group is selected to reflect the make-up of the future population to be tested. The performance of the sample group during the process of standardisation will be used to evaluate individual test scores. During the standardisation process, different kinds of norms can be based. Raw scores of sampled test takers were transformed into standard scores.

Conventionally, result of item analysis based on (CTT) will be employed to pick questions for the final test. The standards used are item difficulty level (the percentage of examinees in the population that responded correctly and their discriminating power) the connection among performance of each learner for each item coupled with performance on the whole. Since CTT cannot sufficiently model responses to each item, Item Response Model is rooted in the assumption that a trait inspires performance, and stipulates how likelihood of responding to a particular items appropriately hinges on the trait assessed, is increasingly used (Burill, 1980).

Test Administration

What is required of a standardised test is consistency of process in their administration. The instructions, materials used and ways of treating queries must be clearly defined. Besides, the conditions the test is conducted in terms of comfort, lighting, well ventilated classroom, absence of interference, and examinee's interest, cooperation, and motivation should be the same for all test takers. Non-conformities in the conduct or in the conditions of testing may influence the interpretation of examinee's.

Test Scoring and aggregation of scores

Detailed step by step guidelines concerning marking of a test must be duly followed. Examiner's discretion thus removed using item selection format – where the test taker is mandatory to choose one precise answer among few option (e.g., objective questions). Tests involving this kind of item may be, and often are, scored by machine, growing the speed and dropping the cost of the operation. Other benefit of selection type items include broad selection of a domain since answering a question takes limited time. Meanwhile, item selection format may however not give a complete assessment of skills and knowledge given in the domain on which the test is centreed, supply- type items may be involved in a test. Such items necessitate the examinee to give an answer (comprising recollection, analysis and synthesis of data), typically in a prose or short written answer. Such items are seen more suitable to produce higher- thinking skills contained in the analysis, synthesis and assessment. Though present criteria to assessment answers will be given, scoring will not be as objective as in the case of choosing type items, giving rise to problems of reliability (Jacob, Darrel and Paulgros, 2018).

Test Score Interpretation

Standardised test are offered with one or two interpretative structures. In a distribution of scores, there is likelihood of tracing the actual position of an examinee's score. When it happens like this, the standard adopted in explaining students' performance is relative, and the score awarded to an examinee is called a norm- reference. A substitute interpretative structure will be given once performance in a test indicates how performance of a testee meets an established standard, criterion, or proficiency level (Jacob, et. al, 2018). Illustratively, if an addition test contained of 60 items selected from all possible items, an examinees percentage – exact mark might be taken to be an estimate of the examinee's intelligence of addition. The percentage of true scores is termed criterion-referenced measure, which is occasionally used to categorise examinee as having attained mastery.

Range of score conversions is given in test manual to enable learner comparisons when norm-referenced test is employed.

They comprise:

i. the proportion of test taker in the sample scoring high or below a given raw score); this is called percentile rank

ii. (direct transformation of 2-scores into random mean score (such as 100) and standard deviation (e.g. 15); called derived or standard score and

iii. Scaled score that reveals a test taker's score relative to the norm groups and the position of the norm group's distribution in relation to that of other group distribution, often done on examinees at a higher score). An illustration of examinees' ability in terms of proficiency level, that combines parts of norm – referencing and criterion – referencing, is increasingly being adopted to present the result of national and international assessment. Separation of performance range into different categories includes scale anchoring which has two constituents: a statistical constituent, that recognizes items that differentiate between consecutive points on the ability scale using specific item characteristics (e.g., proportion of positive answers to items at different score level) and a consensus constituent in which recognized items are used by curriculum experts to provide an explanation of what groups of examinees at, or close to (Jacob, Darrel and Paulgros, 2018).

Test Validity and reliability

Validity and reliability are to be discussed in all assessment activities. Validity is the most important consideration in the use of assessment procedure (Crooks, Kane and Cohen, 1996). According to American Psychological Association (1996), the proposed use of the test determines the extent to which the evidence and the theory support the test scores' interpretation. Part of the test validation process involves providing a conceptual. The concept of construct validity is another important aspect of validity. Examples of construct validity include performance in Reading and Mathematics. It is imperative to cover important aspects of any given construct, while the irrelevances to the construct are jettisoned due to the fact that such irrelevances may affect the test scores in the long run. If the process in the validation is wrong, this may affect the outcomes and may lead to revisiting the tests constructed including the underpinning conceptual framework.

The process of test validity is the collective effort of both the test designer and the test user. Evidence that support test validity include the following:

(i) thorough check of the assessment contents and the construct the test is purported to measure must be done, this is to reveal the relationship between them. The appropriateness of test content is the test specialist's judgment, this is regarded as a form of evidence that should be provided.

(ii) it is also important that a test measure only one construct at a time. Evidence of test unidimensionality will indicate this and it is internal structure evidence based.

(iii) some criteria that the test is expected to predict. This may also include information concerning the relationship among test performances and test developed to evaluate the same domain (evidence of concurrent validity), these are evidence based on relationship of performance on a test to other variables and

(iv) evidence or effects of testing, that is the effect of enlisting students in a special education course or support programmes. The fuse of test scores can be seen to be valid if students benefit for being involved in the educational programme. The influence of additional uses of testing, e.g., increase accountability, also needs to be measured (Jacob, Darrel and Paulgros, 2018).

However, it is a new concept in validity that the consequences of testing should be taken into consideration which is not globally recognised. Before this type

of evidence is considered, there is need to differentiate between the envisioned consequence (such as increased learning outcome in the instructional group that were early assigned based on their performance on a test after a period of time) and unintended consequence. Validity estimation depends on human judgment, though very problematic to achieve. Based on the research studies regarding the determination of some of these criteria and a validity discourse put forward by Cronbach (2000); Crooks, Kane and Cohen (1996) some limitations/threats to the interpretation and use of measurement data for eight components of the measurement procedures surfaced;

i. threat inherent in test administration. an instance is when some students get illicit help, while some are motivated to respond to tasks;
ii. threats inherent in scoring students' test scripts this may happen when scoring rubrics takes cognizance of some indicators of performance, while others are ignored (for example, during verbal language test, vocabulary web is credited, while eloquence or pronunciation is ignored);
iii. threats inherent in aggregating scores of a separate questions to give total or sub-scale scores. For instance, the weights assigned to items during measurement may not reveal the significance of each of the items, though it occurs when differences in score variance for different tasks are not recognized in calculating total scores;
iv. threats inherent in making generalisation from specific item to which accumulative mark is on the entire domain of parallel tasks. If the sample of items from the assessed domain is rather small, it is impossible to extrapolate from the students score to his total score in the assessed domain;
v. threats of generalisation of the domain assessed to a larger domain comprising the entire tasks pertinent in a planned interpretation. If no tasks are included from some significant sections of the target domain (resulting in construct under-representation), it will not be easy to infer from a universe score for the assessed domain to a universe score for the target domain. This will be the case of adequate attention in the assessment is not accorded the content coverage, content quality, and cognitive complexity represented in a curriculum;
vi. threats involved in assessment of the students performance. Unsuitable decisions based on the assessment information might be taken if the assessor

comprehend not the information and restraint arising from its relative nature or the specific procedure adopted to collect it;

vii. threats inherent in the decision or steps to be taken concerning judgments. This happens if the standards adopted in making judgments are unsuitably high or low, if incorrect educational judgments were taken; and if wrong feedback is given to students; and

viii. influence on the students and other members as a result of the measurement procedure, inferences, judgments, and consequences of assessment. This kind of threat occurs if a teacher deserted important curriculum areas, if the teacher came up with wrong expectations for students, low students' motivation, or if pedagogy centreed on the gaining of real knowledge instead of higher level cognitive outcomes;

Reliability on the other hand, refers to internal consistency of measurement of test items on a repeated measure for the same population of examinee (America Psychological Association, 1996). Measurement error is common to the concept of reliability. This is the unsystematic error that occurs due to the fact that a learner is examined with a specific array of questions in a certain situation. This might also be due to variations in scoring essay questions. Systematic error (e.g. error arising due to the fact that a test is much easier than the other and that the two different forms of tests were not adequately matched) cannot be termed measurement error. The hypothetical difference between a learner's true score or total score is called error of measurement. In substantiating evidence to support the reliability of test scores, test users are expected to determine the major sources of error, summary statistics bearing on the size of such errors, and the degree of scores generalisation across different forms, scorers, administrations and additional relevant dimension. The normal deviation of a hypothetical distribution of measurement errors should be provided and reported. It may be based on an internal consistency coefficient an alternate forms coefficient, or a test – retest coefficient (Feldt and Brennan, 1989).

Noticeably, the scores got from a well-constructed test items might not give dependable estimations of individual learner's performance. The separate scores attained by students in simple- based local or international assessment, though appropriate for producing sensibly accurate population estimates, the overall mean scores for male and female students, often cannot may be adopted in reporting different learner performance (the explanation behind is the fact that students may be

examined on a lesser part of the domain of focus, and may not attempt sufficient items to produce dependable estimate of performance across that domain (Jacob, Darrel and Paulgros, 2018).

2.3.7 Standardisation of Performance Test

Standardisation is one of the most tasking aspects of test construction. Osadebe (2001) defined test standardisation as a procedure of generating a valid and reliable test, and the norms establishment. Obtaining a test that is valid and reliable is the first process, and administer to describe test takers in a similar settings or rules. Transforming or norming the raw score of test takers is the next activity using percentile rank, Z-score, T-score and Stannine among others with the knowledge of normal curve. The procedures assist in constructing a standardised test. The test may be used to evaluate students' trait, then contrast the students established on norms. The group through which the test is standardised is norm group. It might be realised using rank percentile, Z-score, Stannine and T-score (Osadebe, 2014). Aiken (1979) points out that the essence of test standardisation is the establishment of norms. The test ought to be conducted with high standard and under same conditions to test takers having the characteristics of the sample to whom the test is meant for the test is to be termed standardised. Observing the array of raw scores in the standardisation group (the norm group) is the essence of the standardisation. The raw scores will therefore be transformed into derived scores or norms such as age equivalent, score equivalent, percentile rank or standard scores. More importantly, examinees can be assessed through relating student's score to the norm table suitable for their peculiar group. In the same manner, norms serve as reference frame for interpreting raw scores (Osadebe, 2014).

In the study of Osadebe (2014), standards were stated through standard and percentile scores (z – score, T – score and Stannine). Worthy of note to say that it is very important having local norms such as norms based on samples of students in particular locality and in a particular school. Then the score of a particular learner can be compared with scores of students across the country, learner in the same locality and students in the same school. Thus, the norms published in test manual are useful for comparing an examinee's score with those a sample of people from various localities, sometimes across section of the nation (Osadebe, 2014). However, pretty often, the test conductor is concerned in investigating how a learner performs

compare with other students within the group rather than a nationally chosen sample. In these situations, the administrator will want to convert the raw scores of the particular school group to local norms. “Local norms are used often for selection and placement purposes in schools or classes”, (Aiken, 1979). It was affirmed “it is imperative to note that standardised tests are important if it is required to juxtapose performances among diverse colleges and levels” (Falaye and Adefisoye, 2016). Norms are very important in that they reveal how others have performed on a test and also enable the comparison of a learner who at any time taken the tests with reference groups or standardised sample.

Ughamadu, Onwegbu and Osunde (1991) upheld that points of Aiken (1979); Gronlund (2006) and Osadebe (2014) type of norms comprise age norms, score norms, standard score norms and percentile norms. The names of the desired score for the score norms, age norms percentile norms and standard norms are score equivalents, age equivalents, and percentile ranks and standard scores respectively. A study carried out on performance test, Okonkwo (2000) employed percentile and standard score norms of nine clearly defined groups. The groups were: rural boys, rural girls, urban boys, urban girls. Rural (boys and girls), urban (boys and girls) boys (rural and urban), girls (rural and urban) and the overall. This is similar to Osadebe (2001). The purpose of many norms is to serve variety of user’s needs. Percentile and standard scores are commonly used in schools. This study had norm group of rural/urban, private/public, male/female pre-schools-children.

Moreover, the normal curve theory assists in the statistical analysis needed for standardisation. Using of normal curve to describe Z-score, T-score, stannine and percentile rank has been widely supported by research findings. These included Aiken (1979); Weis (1999); Frank and Altheon (1994); Owen and Jones (1994); Cohen and Swerdlik (2002); Gronlund (2006) and Osadebe (2014). Once a standardised test items are given to students, their emanating marks shall be normed and contrasted to the preceding related norming group. In this way, the standardised test serves a comparative purpose. When the test is administered to students to judge their performance then it serves assessment purpose. The students’ performance is compared from year to year with their standardisation group (Osadebe, 2014).

2.3.8 Gender and Children's Performance in Reading Literacy

There are research findings which confirmed that boys and girls have differential abilities. For example, Mann, Sassanuma, Sakuma and Mabaki (1990); Lynn (1994) and Ijaiya (2007) found that females performed significantly better on word fluency test, while males achieved significantly higher score in spatial test. Falayajo, Makoju, Okebukola, Onugha and Olubodun (1997) assessed the position of competency of primary four pupils in certain skills: numeracy, literacy and life skills. The national curriculum was used as guideline for drawing up the expected competencies which pupils should have acquired by the end of four years of primary education. These competencies were then translated into test items by a group of classroom teacher and test specialists. Five equivalent form of the test in numeracy and literacy were developed while four were developed for life skills.

A stratified random sample of twenty-four pupils were selected from each of thirty-two schools in each of the thirty states of the federation while twenty-four pupils were selected from each of sixteen schools in the Federal Capital Territory (FCT) of Abuja. The designed sample was twenty-three thousand and forty pupils drawn from nine hundred and sixty schools. The data were analysed using a special item analysis programme for detailed analysis of the test items and the SPSS package. The results revealed that performance in the literacy test was the worst of the three cognitive tests. There was a very little difference in the mean scores of boys and girls, the girls' mean score of 25.8 percent was just a little higher than the boys' mean score of 24.8 percent.

With specific reference to language development, Chandra (2009) reported that despite the variations in sampling and methodology, researchers investigation showed that girls are superior to boys in practically every stage of language development. Bredekamp (1997) added that in general, girls speak earlier than boys and excel in word usage, correctness of sentence structure, and comprehensibility of speech than boys. Similarly, Bagneto and Neisworth (2010) attempted to provide some explanations as to why girls are generally superior to boys in language development using a sociological interactionist approach. According to Bagneto and Neisworth (2010), the superiority of girls is caused partly by their closer association and identification with their mothers. It was further added that, even in infancy, girls seem to gain greater emotional satisfaction from relationships with their mothers and tend to imitate their mother's speech while boys on the contrary wish to identify with

their fathers but they are less successful, probably as fathers are not always at home as mothers. Oedipus Complex of Sigmud Freud that talks about the desire of young children for mutual relationship with opposite sex and committant sense of rivalry contradict the above result. This might probably due to the submission of the brain evidence that states that male fetal brain develops an asymmetry of the two hemispheres with a growth and size of both left and right hemispheres remaining equal while female fetal brain develops an asymmetry in which more rapid and greater neural development of the left hemisphere temporal lobe (that is, language areas) develop.

2.3.9 School Type, Location and Children's Performance in Reading Literacy

The school activities for pre-school children include programmes that could help in making learning a fun and challenging by using children's books and develop cognitively appropriate activities. These activities should seek to stimulate children's natural curiosity; help children learn about the world around them; provide opportunity for children to develop their full potentials and enhance children's literacy skills. There are indications from research finding that structural variables of school such as school type and location could enhance or deter cognitive development of children no matter the pupils' level of development. Literacy – rich environments where pre-school children have access to books and other print materials contribute positively to the child literacy and language development (Hartas, 2011). The Researcher also supports that there are factors influential to reading literacy that is embedded in the school type include teacher's qualification, experience and time devoted for reading activities.

Odinko (2002) found that the more qualified or experienced a teacher is, the more he or she is likely to impact more authentic and reliable facts needed for the intellectual development of the children kept under his/her care. Teacher's qualification could be described along a number of dimensions such as knowledge of the subject matter. Knowledge of pedagogy and the pedagogical content (Goodman, 2009), Odinko (2002) explained the term pedagogical content knowledge to mean knowledge of way of presenting and explaining a subject to make it comprehendible. The qualification of a teacher is also associated with the teacher's capacity to support children learning by providing with the opportunities to work with concrete objects, allow the children to make choices, explore things and idea, experiment and discover

things themselves. Teacher qualification is an important input in teaching/learning situation since quality inputs demand quality outputs. teacher's qualification is associated with the way a teacher plans his lessons, interprets that aims and objectives and delivers the lesson; thus making it possible for teacher's qualification to affect the level of cognitive enrichment bestowed on the children under the care of the teacher.

The differences in academic performance due to location could be as a result of preference by teacher to work in some locations than in others. Johnson (2011) concludes that highly qualified teachers prefer to serve in urban areas than in the rural areas. Many teachers do not accept posting to rural areas and even if they accept the posting, they do not live in those rural areas, thereby not being totally committed to their duties. Many rural schools do not have adequate amenities and facilities (Williams, 2005). There are some locations or cities of Nigeria where there are security challenges, qualified teachers or educational personnel do not go there in search of work or for posting because of the security challenges. This may likely affect the performance of the students in that location (Enu, 2015). According to Jennifer (2009), urban schools excel in the opportunities provided by their location. The researcher also observed that students in the urban have greater access to many educational resources and amenities like electricity, accessible roads to schools and many others, therefore, they have opportunities that are not available for rural students.

2.3.10 Test Development, Standardisation and Item Response Theory Procedure

Empirical studies have been carried out in Item Response Theory and its use in developing and standardising test items. Such studies include Nkpone (2001) who developed and standardised a Physics performance test for senior secondary students, One land two logistics Models and Classical Test Theory (CTT) were used. A total of 2215 samples comprising 866 males and 1349 females who sat for the Senior Secondary School Certificate Examination (SSCE) during May/June 1999 in River state were used. A 60 items multiple test items of Physics performance test was administered to the candidates. Item parameters were estimated using CTT and IRT and the result showed that there was significant relationships among items parameters obtained from one and two logistics model and parameters obtained from CTT.

Igbokwe (2004) used IRT to develop and standardise Mathematics Achievement test items for placement into Federal Unity Schools in Nigeria. The

result of 6,000 pupils who sat for 1998 and 2000 Common Entrance Test for Federal Unity Schools in Nigeria were used. The data was analysed with Bilog-MG software package. The result showed that the average ability of boys was higher than that of girls (male = 1.31; female = 0.10). It was shown that the average ability of pupils from private schools was higher than those from public schools (public = 0.01; private = 1.35). The result showed that pupils who were over 10 years had more Mathematical ability than those who were 10 years or below (≤ 10 years = 0.60; ≥ 11 years = 0.89).

Opasina (2009) used 3 – parameter model of IRT to develop and standardise alternative to practical Physics. The researcher used 11 local government areas of Oyo State, 160 secondary schools and 1545 samples (Boys = 867; Girls = 678). This sample was later divided into Ibadan City (920) and less city (625). Also, 1395 of the sample was selected from public schools while 150 samples were picked from private schools. The items were trial tested with 60 students (boys = 40; Girls = 20). The parameter (difficulty index) ranged between -0.92 and 0.99. It was also revealed that intercorrelation of some sub-groups were high, the correlation values were between -0.48 and 0.63 and standard error of measurement
was from 0.03 to 0.09 which shows a good precision.

Akindele (2004) developed and standardised achievement test items for selection into Nigerian Universities using IRT. The researcher selected 1000 candidates (374 females and 626 males) who sat for 1998 University Matriculation Test conducted by JAMB and used Bilog-MG software package to estimate the parameters and calibrate the items but used SPSS to estimate the means. The study showed that there were significant differences between estimate of parameters using Classical Test Theory and Item Response Theory. The procedure for the estimation of parameters/abilities (Maximum Likelihood Estimate (MLE), marginal a posteriori and Expected a posteriori) did not indicated any significant difference. The estimated talent distribution mean score for females is 0.81, while that of males is 0.94. The result also showed that gender has moderating influence on candidates' academic performance. In comparing the three logistics methods of IRT, it showed that there were significant differences in the values of a, b, and c parameters in one, two and three logistics models. Difference is most noticeable in the estimation of the parameter (difficult level) with mean value of 0.29; 0.56 and 0.97 for one, two and three logistics model respectively. The researcher used 3 parameter model because of

the robustness and quality of estimation of parameters.

Enuwah and Akwa (2014) examined development and standardisation of achievement test in SS Mathematics employing IRT framework. The test captioned, 'Objective Test in Algebraic Processes' (MCTAPIM) was based on SS Mathematics curriculum. One theme under algebraic processes became adequate for the study because IRT model used in the study required a one-dimensional construct. Eleven research questions guided the study. A sample of 1557 students chosen through proportionate strategies random sampling technique was used for the study. An early pool of eighty seven (87) items was constructed, validated repeatedly and trial tested, fifty (50) items survived the item analysis and were used for final administration employing the assistance of Mathematics classroom teacher of each sampled schools. Both trial testing and the final administration were carried out with SS 3 students randomly drawn from Senior Secondary Schools in Cross River state. One hundred (100) students (50) each from both public and private schools) were used for the trail testing, while one thousand five hundred and fifty seven students were used for the final administration.

The research questions were answered using descriptive statistics, while the hypotheses were tested at 0.05 level of significance using the lord's chi-square and learner t-test statistics. All these statistics were generated by IRT software WINSTEPS 372 version. The results from the study showed 47 items from the 50 items of MCTAPIM made up the standardised version of MCTAPIM. Three items were dropped because their outfit mean square values were higher than 1.5 logit units. The instrument was suitable for IRT because IRT assumptions of unidimensionality and local independence of items were satisfied by the MCTAPIM data. Item difficulty parameters fall within -0.55 and 3.13 logit units. The items separated the students into 15.84 strata according to item difficulty parameter. Students' responses were validated while 47 out of 50 item fit the model (Enunwah and Akwa, 2014).

Zanon, Hutz, Yoo and Hambleton (2016) provided a didactic application of IRT and highlighted some of these advantages for psychological test development. IRT was applied to two scales (a positive and a negative affect scale) of a self-report test. Respondents were 853 university students (57% women) between the ages of 17 and 35. IRT analyses revealed that the positive affect scale has items with moderate discrimination and are measuring respondents below the average score more effectively. The negative affect scale also presented items with moderate

discrimination and are evaluating respondents across the trait continuum, however, with much less precision. Some features of IRT were used to show how such results can improve the measurement of the scales. The authors illustrated and emphasised how knowledge of the feature of IRT may allow test makers to require and increase the validity and reliability of other psychological measures.

2.3.11 Validity, Reliability and Utilisation of Instrument of Assessment in Early Childhood Education

This section provides detailed descriptions of tools for assessment in the Early Childhood Education.

Mullen Scales of Early Learning (MSEL) - 1995

These scales were developed by Mullen in the year 1995 for use with children from birth to 68 months (3 years 11 months). The assessment was for the children's answers to questions constructed through the assessor. The test offers full information on a child's cognitive and psycho-motor skills through the use of five scales: Gross motor and four cognitive scales; visual reception, fine motor, expressive language and receptive language. It took around 30 minutes to administer the test. Directives of scoring were also contained within the test administration booklet and scoring should be in a record form. Individual scale gives a raw score which can be compared against age equivalents and the "cognitive" scores can be summarised into an Early Learning Composite (ELC) score. The raw scores can be used in obtaining the child's percentile rank and age equivalent score. The norm group comprised 1849 children, with the exclusion of children with infirmities. Information was collected between 1981 and 1989. A review of cognitive tests by Breadley – Johnson (2001) found limited evidence for concurrent, content and constructive validity of the test.

According to Bishop (2011), the Mullen scale frequently adopted as a measure of cognitive and linguistic skills in research procedures and less commonly mentioned in the general child assessment literature. It is used in research, clinical assessments and longitudinal investigation of children with autistic spectrum disorders. MSEL span the development domains of focus and very easy to score. However, the MSEL has some limitations, including the following:

(i) Specialists required to be greatly competent and be skilled to be able to assess children;

(ii) Normative data, standardised two decades and 3 years back, are now obsolete in comparing with other tests;

(iii) employed in study procedures rather than in overall child assessment;

(iv) there is no evidence concerning its suitability by parents; and

(v) It is relatively costly (Bedford, Walton and Ahn, 2013).

Teacher's Rating of Oral Language and Literacy (TROLL) – 2001

This measure was developed and published by the Centre for the Improvement of Early Reading Performance, University of Michigan School of Education for use with children from 3 to 5 years. The assessment is based on teacher or parent reports on the child skills based on observation. It consists of 25 items which requires the observer to rate the child on a 4-point scale. Information on the child's languages use, reading and writing was provided by the test. It takes 5 -10 minutes to administer the test. Normative sample was based on Northeastern U.S of 900 children. The validity of the instrument is low (below 0.50) but significant. Relationships are reported with direct measures of children's early literacy skills. Moreover, the instrument has high correlation co-efficient (0.80). The instrument is used by teacher to speedily and constantly monitor child's literacy growth and progress. TROLL was conveyed through the publisher for general usage, the early estimates for specificity are good and the completion time is very short.

Phonological Awareness and Literacy Screening Kindergartners (PALS-K) – 2004

The screening instrument is developed and published by the University of Virginia in 2004 for use with individual child in the 5 age range. The assessment is based on the individual child's cognitive ability in the areas of rhyme awareness, beginning sound awareness, alphabet knowledge, letter sounds, spelling, concept of word, word recognition in isolation. The time for the test administration is between 20-25 minutes. The instrument was not normed as reported by the publisher. However, the concurrent validity and reliability were reported to be 0.70 and 0.80 respectively for the test. PALS – K is used for conduct which pupils need instruction in addition to the regular classroom literacy instruction. It is usually adopted as a measure of cognitive skills in research and child assessment literature.

Child Development Inventory (CDI) - 1992

The CDI was developed by Irelon in 1992 for use with children from 15 months to six years. It is designed for the assessment and scoring of children with concerns about development. It involved 300 items with answer sheet attached for parents to fill. There were 270 statements connecting the development of cognitive skills of young children that are noticeable by parents in everyday situations. These items assessed the child's development in eight areas: social, self-help, gross motor, fine motor, expressive language, language comprehensive, letters and numbers. It also includes a General Development Scale and 30 items to identify parent's concerns about their child's health and growth, vision and learning, development and behaviour. Time required to complete the inventory is between 30-50 minutes. For scoring, parents were demanded to respond "Yes or No" to the statements in reporting the activities they have seen their children engaged in. Scoring was done through the frequency count for each scale. Then scores were documented on the CDI profile for comparisons to norms for a child of that age. In the CDI, a child's development was considered to be within the normal range if their scores on the developed cognitive scales are at or above the mean score for children who are 30% younger (that is equivalent to 2 S.D below the mean).

The CDI was developed after thirty years of investigation and scientific knowledge with the Minnesota Child Development Inventory (MCDI). The CDI was standardised on a sample of 568 children, aged one year to six years three months, from Minnesota (Harold and Francis, 1995). The data from the study were not recorded, nevertheless, it was carried out in 1992. The study population was 95% and the authors said "The CDI norms established on the normative sample should not be generalised to groups of children who are different from the norm group. It is better to come up with local norms for particular population or school system". They further stated "the checklist might not be suitable with parents of some ethnic and cultural circles and for parents with less education". According to Child Development Inventory manual, the validity of the CDI was determined in three main ways. First, through test results for norm group, that is, children of younger and older age. Second, by comparing CDI results to psychological test results and by observing the CDI results for children with developmental and other related problems. Moreover, the sample sizes for these studies were relatively small and did not result in figures for sensitivity and specificity. The CDI had been used to follow-up high risk children

(Doig, 1999). Glascoe and Dworkin (1995) studied reasonable quantity of tests depending information given by the parents and based on the CDI being among the performing tests with a sensitivity approximately 80% and sensitivity approaching 90%. The CDI was validated in France with a population sample of 1278 children aged 15-17 months. Sensitivity was 84% and specificity was 92% (Duyne, 2011). The sensitivity and specificity of the CDI were said to be lower for younger children. Rydz (2006) examined certain work in Onebec with recruited 18 months of age children. Of 152 parents sent the CDI, 114 (75%) filled and reverted. The CDI had poor sensitivity (0.50) but has good specificity. CDI covers the development domains of interest and was employed amongst students at high risk for developmental problems. However, the process had poor sensitivity among children in one study.

Bayley Scales of Infant and Toddler Development, Third Edition (Bayley - III) - 2005

The assessment instrument is developed by Bayley (2006) to study every facet of a child's growth. Bayley-III edition covered five development areas. Cognitive, psycho-motor and language are administered on the child (aged three years six months); communication, emotional, social and traits were conducted with questionnaires for parents. Domain subtests can be administered in such a way as to establish level of performance. It took 80 minutes to conduct the test on 10 children aged 13 months and over. Scoring of Bayley – III had been made from the preceding versions. Scoring for every item is either I (credit) or O (no credit). Scores obtainable include draw scores, scaled scores, composite scores, percentile ranks confidence intervals and normative scores. The sample for standardisation especially the cognitive, linguistic and skill scales were used for 2000 US census, and included 1700 children ranges between the ages of 1-42\3 months, further divided into seventeen (17) age groups with 100 children in each category. The sample for the social – emotional scale contained 456 children, while the trait scale had 1,350 children.

The psychometrics centre at the University of Cambridge has carried out work to establish the validity of the norms for use in the United kingdom / Centre for the psychometric, University of Cambridge in the year 2012. A Boyles – third United Kingdom with that of Ireland supplement report gave the results of United Kingdom validation study contained 221 children of 12 to 24 months (Pearson, 2012). The Bayley scales are described as being the most widely used development assessment

scheme. Bayley – III and previous versions have been used with children with range of clinical conditions including prematurity such as the down's condition, pervasive develop cognitive disorder, and linguistic damage. The Baylay scales had been adopted by various nations and the scale is used as a standard against which other tests are compared (Bedford, Walton and Ahn, 2013). Though, Baylay – third edition is not widely adopted as earlier versions (Gollenberg, 2010) and scores seemed to be higher than earlier versions (Bedford, Walton and Ahn, 2013).

2.4 Empirical Review

2.4.1 Relevant Studies on Classical Theory and Item Response Theories

Ogbebor (2016) developed 100 Mock Economics Test and found the items to be unidimensional and locally independent. Classical Test Theory method was used to calibrate all the items in the test, while the 3PL Item Response theory method calibrated only 79 items while 21 items were not calibrated. After calibration with Class Test Theory (CCT), items that were rated as good based on the criterion set were 69 items, while 31 items were poor, while for 3PL-IRT items were selected based on set criterion and how they fit into the Item Characteristic Curve (ICC) model, 50 items were selected, while 29 items were poor. The study found a significant difference in the parameter estimate for CTT and IRT-3PL model. The linear equating method used revealed that Mathematics Equation Test (MET) and 2014 WAEC were equated in their ability scores thus, measuring the same construct.

Enu (2015) developed and validated item banks in Mathematics and Geography for Joint Command Schools Promotion Test (JCSPE) of the Nigeria Army Education Corps with a view to ensuring that the items in the bank are calibrated and of a high quality. The study sampled 600 students in the two subjects (Mathematics and Geography). He adopted the Item Response Theory. The study found that a total of nine, 21 and 38 Mathematics items as well as three, five and four Geography items show DIF in gender, mode of schooling and location in that order. It shows that Mathematics subsets (Algebra, Geometry, Statistics and Probability, Number and Numerations) were difficult with average difficulty levels of 0.90, 0.68, 0.74 and 0.72 respectively and Geography subsets (Regional, Human Economics and Physical) were simple with the average difficulty level of -2.61, -4.80 and -2.38 respectively. The study also found that the difficulty level of Mathematics items was between -1.208

and 2.006 with an average difficulty of 0.16, while the difficulty level of Geography items was between -4.85 and 3.49, with an average of -2.64. The reliability of the Mathematics and the Geography tests, were 0.85 and 0.75 respectively. Mathematics ability of the test takers ranged from -1.32 to 2.45 with a mean of 0.25, whereas, Geography ability of the examinees ranged from -0.53 to 2.45 with a mean of 0.96. It was also found that Mathematics and Geography ICCs were not bell in shape.

Metibemu (2016) examined the comparability of CTT and IRT frameworks in the development of a 50-item multiple-choice Physics Performance Test from a pool of 100-item multiple-choice Physics Performance Test. The study found that the constructed Physics Performance Test was unidimensional and the items stem and options did not give clue for answering another item. Some of the test items Displayed Differential Item Functioning with respect to gender and school location. However, qualitative analysis revealed no incidence of bias in the items. Mantel-Haesznel test - a CTT based method of Differential Item Functioning assessment performed equally as the IRT-based method of DIF assessment. The CTT- based method and IRT- based method of assessing DIF produced similar results in the assessment of DIF in the constructed Physics test items with respect to location. However, the IRT- method out-performed the CTT based-method in assessing DIF with respect to gender. The Classical Test Theory framework deleted more items than the Item Response Theory framework. There was no significant mean difference between the CTT and IRT scoring methods. However, IRT method of scoring produced different test scores for groups of students who had the same score under the CTT scoring method. The IRT and the CTT linear equating methods produced similar results.

Adegoke (2013) compared the effectiveness of Item Response Theory and Classical Test Theory measurement framework in the development of Physics performance test. In the study, 724 Senior Secondary School learners two (age 16-18 years) were randomly sampled and the developed Physics Achievement Test was administered to the students for the final development of the test items. The responses of the students were subjected to item analysis within the Classical Test Theory and Item Response Theory measurement frameworks. The result revealed that 13 items on the basis of the criteria set for difficult indices ($0.30 < p > 0.70$) and 7 items on the basis of discrimination index set ($r_{pbs} < 0.20$) were considered poor. Thus, on the basis of item difficulty and discriminating criteria, 20 items were deleted under Classical

Test Theory item analysis. Under Item Response Theory measurement framework, nine items whose information function did not fall within the targeted information function were considered poor. Also, items with low discriminating indices lower than 0.20 were also considered poor. Under this condition, three items were considered poor. In all, 11 items were deleted. The reliability coefficient of the measurement frameworks (Classical Test Theory and Item Response Theory) were 0.85 and 0.87 respectively. Furthermore, the result revealed comparable evidence when a scatter plot of relationship between r_{pbs} and a, P and b respectively. The findings of the study showed that Item Response Theory and Classical Test Theory item parameters were quite comparable. However, item statistics obtained from Item Response Theory, two parameter model appeared more stable than those from Classical Test Theory, however, for item selection process, Item Response Theory 2-parameter model led to selection of fewer items than Classical Test Theory model.

Ojerinde (2013) applied Classical Test Theory measurement framework and Item Response Theory framework to the item analysis of JAMB pre-test Physics achievement test to empirically compare the effectiveness of the two contrasting measurement in the development of test items. In the study, 50 items of the multiple choice format were administered to 69 students. The responses of the examinees were objected to item analysis within Classical Test Theory framework and the Item response Theory framework. The item analysis within the Item Response Theory framework was executed with the three-parameter Item Response Theory model. The study revealed that under the Classical Test Theory measurement framework, 12 items were considered poor on the basis of the criteria set for the difficulty indices $p> 0.30 > p > 0.70$) and 19 items were considered poor on the basis of discrimination index set ($D \leq 0.20$), while no item was considered poor on basis of error. Under the Item Response Theory, no item was considered poor on the basis of difficulty criteria set ($b < 0.3$). on the basis of criteria set for discrimination indices ($a > 2.95$). Furthermore, the reliability coefficient derived using Classical Test Theory and Item Response Theory revealed 0.49 and 0.67 respectively. Test items were subjected to the two measurement frameworks and conclusion was made on this basis that the item parameters of the measurement frameworks were comparable and that their values can be used almost interchangeably.

Progar and Socan (2008) because of lack of empirical evidence that justifies the theoretical superiority of Item Response Theory over Classical Test Theory in

term of parameters and invariance measures of parameter estimates used the data set from the Thied International Mathematics and Science Study (TIMSS) to address some questions. The findings indicated that the Classical Test Theory and Item Response Theory item/person parameters were very comparable; that the Classical Test Theory and Item Response Theory item parameters show similar invariance property when estimated across different groups of participants, that the Item Response Theory person parameters are more invariant across different item sets, and that the Classical Test Theory item parameters are at least as much invariant in different item sets as the Item Response Theory item parameters. The results further demonstrated that with regards to the invariance property, Item Response Theory item/person parameters are in general empirically superior to Classical Test Theory parameters.

Opasina (2009) developed an alternative to practical Physics multiple choice test. The study determined the psychometric properties of test items using Item Response Theory. The unidimensionality of the developed items was established. Items whose difficulty index ranges from -1 to 1, discrimination index ranges from 0.15 to 3.0 and the vulnerability to guessing did not exceed 0.35 were all selected for the final Physics practical multiple choice test.

Wiberg (2004) evaluated the item parameters of Swedish driving-license test via Classical Test Theory and Item Response Theory measurement frameworks, with the use of 5,404 test takers who took the multiple choice version of the test that consisted of 65 items. All the test takers answered each of the 65 regular items in the test. Among the test takers, 43.4% were females and 56.4% were males and their average age was 26 years. The study examined the three Item Response Theory models with responses of the test takers to choose which of the Item Response Theory model best fit the data. The three-parameter Item Response Theory model came out the best for the assessment of the test result. The item parameters obtained using the 3-parameter model was then compared with the item parameters obtained from the analysis of test items under the Classical Test Theory. The findings showed that the item parameters obtained from the 3-parameter Item Response Theory model and Classical Test Theory were highly related. However, the study submitted that Item Response Theory provided a measure of extent of guessing associated with each item of the test.

Nenty (2004) reviewed CTT and IRT and compared the two theories. The study points to the fact that CTT has sustained educational measurement for almost a century. Its measurement produced results that are at best meaningful only in extremely limited situations. It is believed from the review that IRT has arrived on time to take educational measurement into a century filled with bright hopes and an exciting future, since IRT provides opportunity to yield invariant item and latent trait estimates, standard errors conditional on trait level, and trait estimates anchored on item content. Item response theory can also facilitate the assessment of Differential Item Functioning, inclusion of items with different- response formats in the same scale and assessment of person fit.

Zumbo and Hubley (2017) made a report on introduction of objectivity into educational measurement through the use of IRT. The report states that CTT lacks the property of objectivity. This was said to be because the item difficulty and the estimates of a person's ability depend on the characteristics of the measurement instrument as well as groups with which the examinee took the test. Item response theory is said to provide a tremendous improvement over CTT because its item parameter and person ability estimates are sample free which makes educational and psychological measurement to be objective.

Akindele (2004) developed a prototype item bank for the Nigerian JAMB English Language test items using IRT. The items stored in the bank have an outlined curriculum strand represented by the items. The item bank also contains test item parameters, item statistics and calibration information which will facilitate important decision making in test development. The JAMB item bank will make it easier to create a parallel test or a tailored test with pre-specified item characteristic. In the study, analysis of data set was accomplished through the use of SPSS and item calibration was executed by BILOG MG. The results for the item analysis for CTT and IRT were presented. The study revealed that the scale scores of the three subtests contained in the test did not reveal any significant difference.

Courville (2005) replicated Fan (1998) study with the responses of 80,000 of which 40,000 were male and 40,000 were female to the ACT assessment. Test composed of four tests: English, Mathematics, Reading and Science. These sub-tests were of the multiple-choice type response format. The findings indicated that in a variety of conditions, the two measurement frameworks produced similar item and person statistics. Also, Classical Test Theory based parameters were found to be as

invariant as the Item Response Theory based item parameters. Furthermore, Stage (2003) compared the behaviour of item and person parameter estimates of the Classical Test Theory and Item Response Theory based measurement framework in order to decide whether Classical Test Theory which has been the framework used for the development and assembly of the Swedish Scholastic Intelligence Test (SWESAT) should be replaced by Item Response Theory. The study consisted of 2,461 test takers (1,349 females and 1,112 males) that were randomly selected from 82,506 examinees who took part in the SWESAT in spring 1996. The results of these examinees were the data used for the study. In order to control for the effect of the theoretical framework on which the test was developed and assembled the analysis of the pre-test (the trial testing) and the regular test (the real test) were done under the Classical Test Theory and Item Response Theory measurement.

This study also established which of the IRT Model for dichotomous data fits the Draft Early Reading Literacy Test (ERLT) for pre-school children, J. Metrik Version 4.1.1 was used to establish this and the result indicated that 2 parameter Logistic Model is most appropriate for the DRAFT-ERLT data. The DRAFT 226 Early Reading Literacy Test for pre-school children's item parameters were established using IRT framework. The item discrimination (a) values ranged between 0.12 and 6.36 while that of difficulty index was between 0.85 and 7.92. The researcher also selected items from the 226 ERLT using IRT framework, 163 items were selected using the difficulty range of -3 to +3. The study also established the ability score of 776 examinees that responded to the 163 DEV-ERLT, the T-score showed minimum (19.80), maximum (71.70) and mean (51.10). The estimates reliability for sub-sets DEV-ERLT that the reliability coefficient ranged between 0.62 (identification of colours) and 0.99 (identification of letters).

2.4.2 Relevant Studies on Standardised Testing and Academic Performance

A number of studies have been conducted on the impact of standardised tests on learners' academic performance. Morgan (2014) investigated the effect of the Score Four Literacy Test (G4LT) in Jamaica as a preparatory agent for improved performances in the literacy level of pupils who completed the standardised Score Six Performance Test (GSAT). The research was based on the use of quantitative analysis and variable correlations. Data was collected from a secondary source on performance rate of pupils who completed the G4LT and the GSAT in the areas of

Language Arts and Communication Task from 2003 to 2013. The results indicated that indeed, there exists a relationship scheme between the performance levels on the G4LT and the GSAT tests.

Kennedy (2003) reviewed the nature of standardised test preparation for schools, using the following: a model and adoption of a systematic approach to improvement and change aligning the educational process with desired outcomes, aligning the educational process with imputes, creating positive change and also undertook a case study of an elementary school to test his theory. It was found that standardised tests had the potential to improve schools and also to do a great deal of damage to the teaching and learning process.

Heckman, Pinto and Savelyer (2013) suggested using measures of behaviour such as participation in learner activities and other observations by teacher and school administrators to complement standardised performance tests when evaluating students and schools. Several recent studies demonstrated that this is a promising approach. Heckman, Pinto and Savelyer (2013) showed that teacher ratings of elementary school children's behaviours are strong predictors of adult outcomes and that early childhood interventions promoted the non-cognitive skills measured by these ratings. Heckman et. al. (2013) studied the effect of teacher on learner cognitive and non-cognitive skills. He measured cognitive skills using standardised performance test scores, while measuring non-cognitive skills using absences, suspensions, scores and score progression. These measures of non-cognitive skills predict adult outcomes with a strength similar to measures of cognitive ability. His measures of non-cognitive skills are commonly available from the administrative records of schools.

Cosgrove, Shiel, Sofroniou, Zastrutzki and Short (2005) found that performance of students on Programme for International Learner Assessment Mathematics has relationship on their performance in the JSS Mathematics Test done in either 2002 or 2003 in Ireland. However, the relationship amongst the performances in Mathematics and overall performance on PISA Mathematics was found to be 0.75. Correlation between Junior Secondary School Certificate Test performance and PISA Mathematics content areas ranged from 0.68 (space and shape) to 0.74 (uncertainty). Similar results were obtained when performance on the Junior Certificate Test in science was correlated with performance on PISA 2006 Science (r = 0.70). A similar correlation (0.68) was found between Third International

Mathematics and Science Study in year 1995. Mathematics scores in first year post – primary education and performance on the Junior Certificate Mathematics test at the end of third year, though in this case, there was a two- year time lag between the two assessments (Sofronion and Kellaghan, 2004). In United Kingdom, significant correlation coefficients were found between students' key stage 3 level scores in English (at age 14) and PISA ($r = 0.73$). Relationship was rather stronger for PISA Mathematics ($r = 0.82$) and PISA Science ($r = 0.83$). In Iceland, a relationship of 0.60 was got between performance on Icelandic Linguistic Test and PISA reading literacy.

Studies have also identified issues related to performance that could be transmitted from one educational setting to another. For example, studies conducted by the National Centre for Research on Assessment, Standards and Learner Testing (CRESST) on the impact on students language background and their performance indicated that learner language background affects students performance in content-based areas such as Mathematics and Science (Abedi and Lord, 2001; Abedi, Leon and Mirocha, 2001) Abedi (2002) using data from several locations across the U.S., examined the impact of students language background on the outcome of achievement test. The result of the analyses indicated that students assessment results might be confounded by their language background variables. English Language Students (ELLs) generally performed lower than non-ELL students on Reading, Science and Mathematics. This showed a strong indication of the impact of language proficiency on assessment.

Moreover, the level of impact of language proficiency assessment of ELL students is greater in the content areas with higher language demand. For instance, analyses showed that ELL and non-ELL students had the greatest performance difference in the language – related sub-scales of tests in areas such as Reading. The gap between the performances of ELL and non-ELL students was smaller in Science and virtually non-existent in Mathematics computation subscale, where language presumably had the least impact on item comprehension. Nonetheless, the relationship among standardised performance test scores and external criterion measures was significantly larger for non- ELL students than for the ELL students. Analysis of the structural relationships between individual items and between items and the total test scores showed a major difference between ELL and non-ELL students.

White (2007) sought to address the question: Are girls better reader than boys through the use of Ontario's 2002 large-scale administration of the Ontario's

Secondary School Literacy Test (OSSLT). It was concluded that the notation of under-performance of boys in the area of reading performance has been greatly overstated. National Reading Panel (2009), in gender differences in spelling performance in Score 1 through 6, compared the performance of boys and girls on standardised evidence reading and written spellings tests for students from high, medium and low achieving school in six geographical areas of United States. It was found that girls scored significantly higher, resulting in the conclusion that girls spell better than boys at all levels.

Osadebe (2014) Standardised Economic Performance Test for senior secondary school students in Nigeria. Three research questions were raised for the study. The standardised test in Economics was first constructed by an expert as a valid and reliable test. The test was then used for standardisation in the study. It was administered to 3,000 students using the same guidelines with no case of malpractice, the sex, location and school- type norm of students were considered for standardisation. The measurement of students in form of percentile rank, Z- score, T-Score and Stannine statistics were used as derived scores to normalise the students' raw scores, using the knowledge of normal curve as the theoretical base. The result showed that the test scores were normally distributed with the normal curve for the sex, location with type of school norm of students by means of percentile rank, Z-Score, T-Score and Stannine. Using normal curve to describe Z-score, T-score, Stannine and Percentile rank has been widely supported by research findings. These included Aiken (1979), Nunnally and Bernstein (1994), Owen and Jones (1994), Frank and Altheon (1994), Weis (1999), Cohen and Swerdlik (2002), Gronlund (2006), Osadebe (2014). Once standardised instruments are subsequently given to students, their scores are normed and compared with the previous related norming group. In this way, the standardised test serves a relative purpose. When the test is administered to students to judge their performance, then, it serves as assessment purpose. The students' performance could be compared from year to year with their standardisation group (Osadebe, 2014).

This study also standardised Early Reading Literacy Test (ERLT) for pre-school children in Nigeria. Eight research questions were answered in the study. The researcher developed 226 DRAFT-ERLT which were pilot tested and 226 items survived of which the item parameters were established using IRT framework - the 226 DRAFT-ERLT items were also subjected to the test of Differential Item

Functioning. From the 226 DRAFT-ERLT items, 163 were selected after the validation, reliability and parameters of those items were established. The 163 DEV-ERLT were administered on 776 pre-school children following the condition of standardisation of instrument, scores of 776 pre-schoolers were normed based on the age, gender, type of school and location of the examinees using T-score, percentile and stannine.

2.5 Appraisal of Literature

Various studies in Nigeria and other climes have been carried out on development, validation and standardisation of assessment instruments using Item Response Theory framework. However, there are several gaps observed in each of these studies. For instance, a study focused on development of alternative to practical Physics multiple choice tests. The researcher determined the psychometric properties of the test using Item Response Theory Framework. This is developed for Senior Secondary Students not pre-school children and the instrument was not standardised. Another research was on standardisation of Economics achievement instrument for Senior Secondary School Students in Nigeria. The test scores were distributed with normal curve for the sex, location and school type norms adopting percentile rank, z-score, T-score and Stannine. This work is not relevant to pre-school children. Also, Mathematics Performance Test items were developed for placement into the Federal Unity Schools in Nigeria. The work focused on determining the psychometric properties of the test items using Item Response Theory framework. The validated test items are not relevant to pre-school children. Besides, the researcher did not attempt to standardise the test items. Physics Performance Test was also developed for Senior Secondary School students, using Item Response Theory and Classical Test Theory framework. Although, the unidimensionality of developed test items was established, local independence of the test items was not established. Besides, the test items are not useful to pre-school children. Another Researcher had developed and validated item banks in Mathematics and Geography for joint command schools promotion test of the Army Education Corps with a view to ensure that the item in the bank are calibrated and of high quality. The study focused on Senior Secondary students not on pre-primary children. Other researchers had also focused on the Differential Item Functioning and the establishment of item statistics of the test for selection into Nigeria Universities. Other works concentrated on the establishment of

psychometric properties of Physics Test Items obtained from IRT and CTT. In the above mentioned studies, those instruments were not relevant to pre-school level.

In other climes, Item Response Theory framework has been applied in the validation of test items. Some Researchers provided a didactic application of IRT and highlighted some of its advantages for psychological test development. IRT was applied to two scales (a positive and a negative effect scales) of self- report test. The authors illustrated and emphasised how knowledge of the feature of IRT may allow test makers to inquire and increase the validity and reliability of other psychological measures, however, this has not been fully established at the preschool level.

Specifically, numerous assessment instruments have been developed and validated and standardised in other climes for pre-school children. For example, Muller scale of Early learning, was developed for children from birth to 3 years 11 months. Teachers Rating of Oral Language and Literary was developed for children from 3 to 5 years. Phonological Awareness and Literacy Screening Kindergarteners was developed for children in the 5 years age range. IRT is suitable for these instruments since IRT assumptions of unidimensionality and local independence of items are satisfied by software WINSTEPS 372 version. However, these instruments are not culturally relevant to Nigerian children. More importantly, many of these instruments on Early Reading Literacy skills developed in other climes were assessing two and at most three reading sub-skills, while this study developed and standardised instrument on early reading literacy covering five major sub-skills in reading literacy skills,

From the foregoing, numerous studies have focused more on the development, validation and standardisation of test items for higher school students and not for pre-school children in Nigeria. This study therefore, developed and standardised culturally appropriate Early Reading Literacy Test for pre-school children. The test scores obtained were transformed to percentile rank and normed according to the gender, age, type of school and the location of schools attended by the examinees.

CHAPTER THREE
METHODOLOGY

This chapter describes the research methodology. It includes research type, population, sampling technique and sample, instrumentation, method of data collection and procedure for data analysis.

3.1 Research Design

This study is an instrumentation research that adopted counter-balance design. Counter-balance is a design that allows possible order of administering the instrument so that none of the items suffered.(Niel, Tim and Daniel, 2010). It also assisted in reducing errors that might emanate from test administration such as random and systematic errors which might unknowingly be taken as variance.

Table 3.1: Order of administration of the instrument using counter-balance design

Item Section	B and C	D and E	F and G	H and I
Group	1	2	3	4
Group	2	3	4	1
Group	3	4	1	2
Group	4	1	2	3

While participants in group one was answering sections Band C, participants in group two was answering sections D and E, participants in group three was answering sections F and G and participants in group four was answering sections H and I and so on.

3.2 Population

The target population for this study comprised all pre-school children from public and private schools in Oyo State that are in final pre-school classes.

3.3 Sampling Technique and Sample

Sampling was carried out in two stages. Stage 1 was for item calibration and parameter estimation, while Stage 2 was for standardisation of Early Reading Literacy Test.

3.3.1 Stage I: Sampling and sample for Item calibration and parameter estimation

Multi-stage sampling procedure was employed for this stage. From the existing three (3) senatorial districts in Oyo State, namely; Oyo Central, Oyo South and Oyo North. Proportionate to size sample technique was used to select four (4) Local Government Areas in Oyo Central, four (4) from Oyo North and three (3) from Oyo South, making eleven (11) LGAs altogether. Also, in the sampled LGAs, proportionate to size sampling technique was used to select fourteen (14) schools (7 public and 7 private) from Oyo Central, ten (10) schools (six public and four private) from Oyo North and sixteen (16) schools (six public and ten private) from Oyo South senatorial district. In each of the sampled schools, where there were more than one final pre-school classes, an intact class of the final level of pre-school children was randomly selected. In a situation where only one existed, such a class was automatically selected to participate in the study. Thus, forty (40) pre-schools (19 public and 21 private) intact classes were involved in this study. The number of pre-school children sampled was five hundred and sixty (560) - 220 males and 340 females

Table 3.1.1: Sample frame for item calibration and parameter estimation

Senatorial District	No of LGAs	No. of LGAs Selected (1/3)	No. of Public Schools in the Selected LGAs	No. of Private Schools in the Selected LGAs	No. of Public School Selected (3%)	No. of private School Selected (3%)	No. of preschool children selected	Total Number by Gender	
								M	F
Oyo Central	11	4	212	229	7	7	196		
Oyo North	13	4	205	138	6	4	140		
Oyo South	9	3	206	344	6	10	224		
Total	33	11	623	711	19	21	560	220	340
								560	

3.3.2 Stage II: Sampling and sample for Standardisation

From the existing three (3) senatorial districts in Oyo State, proportionate to size sampling technique was used to select 4 local governments from Oyo Central in which 26 schools (8 public and 18 private) were selected, 4 Local Governments from Oyo North in which 10 schools (7 public and 3 private) were selected and three (3) Local Governments from Oyo South in which 19 schools (13 private and 6 public) were selected respectively. Thus, in all, eleven (11) Local Government Areas (LGAs), 55 schools and 55 intact classes were involved in the study. The number of children that participated were seven hundred and seventy-six (776).(368 males and 408 females)

Table 3.2: Sample frame selected for Standardisation

Senatorial District	No of LGAs	No. of LGAs Selected (1/3)	No. of Public Schools in the Selected LGAs	No. of Private Schools in the Selected LGAs	No. of Public School Selected (3%)	No. of private School Selected (3%)	No. of pre-school children selected	Total Number by Gender	
								M	F
Oyo Central	11	4	258	594	8	18	368		
Oyo North	13	4	224	92	7	3	142		
Oyo South	09	3	205	423	6	13	266		
Total	33	11	687	1,109	21	34	776	368	408
								776	

3.4 Instrumentation

3.4.1 Early Reading Literacy Test (ERLT)

This instrument "Early Reading Literacy Test" (ERLT) was constructed by the researcher to test the major pre-reading skills in children that attend pre-schools. It has nine sections (A - H). Section A generated the demographic information of the pre-school children (such as gender, name and age) and that of the centre which they attend (type and location). This information was written by the researcher for each child. Section B – H consist of 226 test items generated in accordance with the Integrated Early Childhood Care and Development Education (IECCDE) curriculum, one-year compulsory pre - primary education curriculum and Nigeria Early Learning Development Standards compiled by Nigerian Educational Research and Development Council (NERDC). Some textbooks on emergent reading approved by the Ministry of Education for nursery education also served as guide in constructing

the test. The instrument consists of items that captured major pre- skills in early reading literacy skills; knowledge about print, recognition of signs and symbols and identification of (letters, objects, parts of the human body. animals, birds and human beings) as outlined in One-year Pre-Primary School Education Curriculum (FRN 2014, page 15 – 16 and Early Learning Development Standards for Nigeria page 37 (NERDC 2014).

Table 3.3: Sub- skills in Early Reading Skills as captured in the DRAFT ERLT

Sub-Skills in Early Reading Literacy skills	Sub-Skills as captured in the instrument	Pages addressing each sub-skills in the instrument
Oral language ability	Identification of colours, shapes, fruits, animals etc	Pgs, 4, 5, 6. 7, 8, 9, 10, 11, 12, 13, 14, 15 and 16
Phonological awareness	Identification of alphabets and ability to sing rhymes correctly	Pgs 1, 2, 3, 22 and 23
Reading Fluency	Reading fluency	Pgs. 18, 19, 20 and 21
Reading Comprehension	Picture Reading	Pgs, 24 and 25
Knowledge about print	Recognition of Signs and Symbols	Pg. 17

3.4.1.1 Development of Early Reading Literacy Test (ERLT)

The purpose of Early Reading Literacy Test (ERLT) is to measure the major pre-reading skills in children that attend pre-schools, as spelt out in the Nigeria Early Learning Development Standards compiled by Nigerian Educational Research and Development Council (NERDC). The test may be used for placement purpose in primary education and also for identification of aspects of pre-reading skills where children are having difficulty. It can also be useful in comparing children's skills in reading literacy across the age, gender, type of school and location of the schools they attend so as to device better means of stimulating children to learn better on those aspects.

The test was administered on children in final pre-school level both in public and private primary schools in Nigeria. ERLT is culturally appropriate to children in Nigeria. Since the test is designed for children, it is pictorial and colourful, which made it attractive and assisted in sustaining their interests and attention span. The following steps were taken in the development of the instrument.

According to American Psychological Association after, defining the purpose and use of the test is the development of test, the next action is defining of the format. That is, how the test will be presented and how the test takers will respond. ERLT was administered orally. The researcher presented the test before each child, asked questions on each item one after the other and recorded the child's response immediately on the score sheet. The test blueprint developed on ERLT covered major pre-skills in reading according to the approved one-year compulsory pre - primary education curriculum and Nigeria Early Learning Development Standards approved by the Federal Ministry of Education. It included knowledge about print, identification of signs, symbols, objects, animals, shapes and colours, ability to sing rhymes and read pictures. ERLT provides a good representation of the curriculum.

In the development of the experimental test form, the researcher consulted approved curriculum, early learning standards, approved texts on rhymes, stories and starting reading, while drafting the items. Initially, the researcher drafted 255 items. These items were subjected to face and content validation by the researcher's supervisor, four experienced caregivers in private and public pre-schools and some lecturers in the Institute of Education. Many of the items were re- drafted while some were discarded due to one inadequacy or the other. The researcher re-drafted the items leaving 242 items that were more attractive, unambiguous, clear and colourful pictures that tested one's pre-skill at a time. The face and content validity of the items were looked into by the researcher's supervisor, five experts in Educational Evaluation, two experts in Early Childhood Care and Development Education (ECCDE) and three experienced caregivers. The researcher was advised to visit some ECCDE Centres and find out their correct operational vocabulary and language of instructions. This led to adjustment of some of the instructions meant to guide the researcher/ research assistants in administering the test. In pilot testing of the instrument, 242 items were administered on 45 pre-school children from two pre-schools in Osun Central Local Government, Osogbo at Osun state to establish the valid items. Empirical Reliability- an output of IRT was used to establish the reliability of the instrument and it was 0.84 index.

The researcher assembled the selected experimental test items. After removing, and editing of the items where necessary, the researcher was left with 226 items. This large numbers of items was necessary since the test is to be subjected to calibration and standardisation. The test is meant to be produced and used in Nigeria.

Test users may decide to take any section of the test at a particular time for a specific purpose. The researcher also developed ERLT scoring sheet, guidelines on the administration of ERLT and ERLT review at a glance which showed some information like year of publication, total number of sample, Stannine points, Differential Item Functioning and so on at the end of the test standardisation. Instrument was tried out, validated, and assembled as the final form of the test that was administered on 560 sample, the instrument was calibrated and 163 items survived. The items and persons statistics were estimated. This was administered on a larger sample of 776 test takers, after which normative data were generated.

The test items covered the lower order of Bloom's taxonomy in the cognitive domain as it applies to each item indicated in Table 3.4.

Table 3.4: DRAFT Early Reading Literacy Test Blueprint

Content	Knowledge	%	Comprehension	%	Application	%	Total	%
Letter Identification	Item no. (1–104) = 104 Items	45.6			-		104	45.6
Objects Identification	Item no. (105-164) = 59 Items	25.9			-		59	25.9
Identification of Shapes	Item No. (165-169) = 5 Items	2.2			-		5	2.2
Identification of Colours	Item No. (170-173) = 4 Items	1.8			-		4	1.8
Rhymes			Item No. (198-206) = 9 Items	3.9	-		9	3.9
Reading Fluency			Item No. (178-197)= 23 Items	10.0	-		20	10.0
Recognition of Signs and Symbols					Item No (174–177) = 4 Items	1.8	4	1.8
Picture Reading					Item No. (207 – 226) = 20 items	8.8	20	8.8
Total	**172**	**75.5**	**29**	**13.9**	**24**	**10.6**	**226**	**100**

Identifications of letters, things, objects and people carried larger percentage in the instrument because learning to read and write starts with learning the code which is the alphabets and phonological awareness. Also, there are 26 alphabets in English language in which the capital, small letters in a direct and juggled forms were tested.

Table 3.5 shows the test blueprint for the DEV - ERLT after some sub-skills and items that did not survive were deleted from the instrument.

Table 3.5: Developed ERLT Test Blue Print after Validation

Content	Knowledge	%	Comprehension	%	Application	%	Total	%
Letter identification	Item No. (1-82) = 82 items	50.3					82	50.3
Objects identification	Item No. (83-127) = 45 items	27.6					45	27.6
Identification of colour	Item No. (128-129) = 2 items	1.2					2	1.2
Recognition of signs and symbols	Item No. (130-132) = 3 items	1.8					3	1.8
Reading fluency			Item No. (133-143) = 11 items	6.8			11	6.8
Rhymes			Item No (144-148) = 5 items	3.1			5	3.1
Picture reading	-				Item No. (149-163) = 15 items	9.2	15	9.2
	132	**80.9**	**16**	**9.9**	**15**	**9.2**	**163**	**100**

3.4.1.2 Manual on Administration of Early Reading Literacy Test (ERLT)

Early Reading Literacy Test (ERLT) was developed to test the major pre-reading skills of children that attend pre-schools in Nigeria as spelt out in the Nigeria Early Learning Development Standards and one-year compulsory pre-school education curriculum compiled by Nigerian Educational Research and Development Council (NERDC). ERLT is meant for children at the final pre-school classes. The test was culturally appropriate to Nigerian children. ERLT is in two sections: Section A generated demographic information such as age, gender and name of pre-school

child. Information on the centre the child attends like the type of school and the school location were also included. Section B – H consist of 226 test items on: Letter identification, identification of objects, part of human body, fruits, animals, identification of colours, recognition of signs and symbols, ability to sing rhyme, reading fluency and picture reading. ERLT should be administered by: experienced pre-primary class teachers; Early Childhood care and Development Education specialists; Educational Evaluators; Child Psychologists and research assistants that have been trained on the purpose and the administration of the test.

The test administrator should ensure that testing security and integrity is maintained. There should be no leakage of the test. No reproduction of the test in any form should be allowed. Since it would be copyrighted. The test administrators need to familiarise themselves with the children because children accept and relate freely with persons they are familiar with, not strangers. They should relate with the children, be part of the activities in the centres before the day of the test's administration. This calls for a pre-visit to the centres for external administrator, test administrator should also be familiar with the test, the materials and as well master the procedures before the testing day. Testing should be scheduled at a time that facilitates test takers' maximum performance. Preferably in the morning, before 12.00 noon. Test takers should be allowed to rest in between the test sections if need be, considering the children's attention span.

No time limit should be set, every child must be allowed to move at his/her own pace. However, the test administrator should note the average time spent on the testing. The sitting arrangement should be in a way that one will not distract another. The centre/classroom should remain the normal way it used to be, no need to remove displayed charts, toys and the rest. There should not be distraction around and within the venue. Each section of the test carries an instruction. The test administrator should give the instruction in his/her natural tone. The instruction may be repeated, if necessary. The test must be administered through one-on-one interaction with the test takers. It may be necessary to employ the service of more test administrators who should have been trained depending on the number of test takers. The test administrator should interact with a child at a time, and read out the instruction and allow each child to attempt all question. The test should be administered orally and the administrator should record the score immediately in the score sheet.

Details of the manual on the administration of the instrument can be found on appendix II.

3.4.2 Validation of the instrument

Two specialists in Early Childhood Care and Development Education (ECCDE), three experienced ECCDE caregivers, two specialists in English Language Studies and two Educational Evaluators were consulted for face scrutiny and content validation.

The 242 items were pilot tested on 45 pre-school children selected from two pre-schools in Osun Central, Osogbo in Osun state which is different from the sample used in the main study. Empirical reliability - an output of IRT was used to establish the reliability of the items, the value derived was 0.84 index. Thereafter, the survived items of 226 were administered on 560 final pre-school children in selected 40 pre-schools of intact classes for item calibration and selection of valid items. Also, to determine the reliability of the instrument, the internal consistency was investigated using IRT framework - empirical reliability which ranged from 0.62 to 0.99. 163 items survived and these were administered on a larger sample of 776 examinees for usage and assessment. The examinees' ability is used to categorized them and their scores were normed.

3.5 Scoring and Item Selection

The items were scored manually. Each item carries equal weight and mark (that is (1) each). Item selection in IRT model was based on the intended purpose of the test and an acceptable range of difficulty level of -3 to +3. Items that met the difficulty range of -3 to +3 were rated as survived items (Baker, 2001). 65 items fell out of the range. Though, some items might not fall within the stipulated statistical range, it does not mean the items are bad. The implication is that teachers should pay more attention to those items by repeating the stimulation several times until the preschoolers get them right. Therefore, items that were found playing a vital role in the subtest and are contextual valid were retained in the study. This study operationalised the items as detailed in 1.6.2 of the work.

3.6 Procedures for Data Collection

The researcher collected a letter of introduction from International Centre for Educational Evaluation. (ICEE) to Oyo State Ministry of Education, State Universal Basic Education Board and Head teachers/Proprietors of the centres involved in the study. The head teachers of the schools were briefed on the purpose of the study. Fifteen Research Assistants who were co - opted from the educational evaluation and Early Childhood Education fields of study were trained by the researcher on the purpose of the exercise, the administration of the instrument and scoring of the participants' responses. Thereafter, practical sessions of three days were held and necessary corrections were made. Visitation to the Early Childhood Care and Development centres involved in the study was done twice. The first visit was a familiarisation tour. At the first touring, the research assistants helped the caregivers with the centres routines by attending to the children's needs. The familiarisation tour was necessary because of the peculiarity of the samples of the study. Children respond better and interact freely with people they are familiar with, not strangers. This assisted in making the research assistants acceptable to children on the real day of administration of the test, virtually all the children were happy and cooperated. The participating centres were not informed of the exact day of the next visit, this assisted in eliciting response from the participants in the natural settings and tones of the centres.

Intact class of final level of pre – school children was the sample of the study. Twelve research assistants worked with the researcher which made the administration of the test a bit faster considering the children's attention span and the early scheduled activities in the centres. The caregivers on ground at the centre assisted only in setting suitable environment for the administration of the test. They made the children relaxed, ready and cooperate with the researcher/research assistants. The instrument was administered on the children one on one. Each section of the instrument has instructions at the top in a language that children are familiar with. The researcher/research assistants followed the guidelines in the test administration booklet by asking appropriate questions and recorded the child's scores immediately on the score sheet - '1' for correct response or '0' for unanswered or not rightly answered item. The average time spent by the test takers in responding to the whole instrument was 15 minutes. The examinees were grouped into four(1, 2, 3 and 4) and the test was also divided into four sections (B and C, D and E, F and G and H and I).

While sections B and C were administered on group 1; D and E were administered on group 2; F and G were administered on group 3 and H and I on group 4. The administration of the sections of the test was interchanged among the groups. This was aimed at giving each section of the instrument equal chance of being attempted by the test takers at a given good level of attention and interest.

3.7 Procedure of Administration

The researcher/research assistants administered the test, following these rules:

(1) researcher/research assistants administered the test through one-on-one interaction;

(2) each child was allowed to attempt the questions one-on-one.;

(3) no child was rushed. Thus, each child was given enough time to respond;

(4) in the aspect of song/rhyme, the researcher/research assistants led the song/rhyme while the child followed and completed it; and

(5) as the child responded, the researcher/ research assistants/ticked appropriately on the score sheet.

3.8 Assumption of the Item Response Theory

To assess dimensionality of test information, two types are evident. Whenever only a single score is reported for a test, there is an implicit assumption that the items share a common primary construct, it means unidimensionality is tenable, otherwise multidimensionality is evident. For this study, Stout's test of essential unidimensionality, implemented in DIMTEST version 1.0, was used to establish assumption of dimensionality of the test information. To perform the test, items were divided into two subtests that are as dimensionally distinct as possible. The first of these subtests, called the assessment subtest (AT), is chosen such that the items contained within the partition are homogenous to one another, but as dimensionally distinct from the remaining items as possible. The second subtest is called partitioning subtest (PT), which consists of all items not used in assessment subtest (AT). More importantly, items that might form a secondary dimension, the Assessment Subtests were selected empirically, using the HCA/CCPROX cluster procedure and DETECT statistic in DIMTEST. The examinee cluster was tested to see if it was dimensionally distinct from the remainder of the test. A random sample of 30% of the examinees

was used to select the Assessment Subtest, and the remaining sample was used for the dimensionality test (Demars, 2010). The null and alternative hypotheses tested by DIMTEST are given by (Stout *et al.,* 1996). They are:

H_0: $AT \cup PT$ satisfies essential unidimensionality ($d = 1$)

H_i: $AT \cup PT$ fails to satisfy $d = 1$

The null hypothesis posits that the AT and PT partitions assess the same dominant underlying dimension, while the alternative hypothesis implies that the items in the AT partition are best represented by a dimension that is distinct from that driving responses to the PT items. Also, after correction for bias, the test statistic, T, is assumed normally distributed. The null hypothesis for T is that the responses are unidimensional, thus, failure to reject the null hypothesis signifies that the assumption of unidimensionality is justifiable. If otherwise, multidimensionality is manifest (Demars, 2010).Table below presents DIMTEST statistics of unidimensionality of the 226 items contained in the ERLT performance.

Table 3.6: Dimtest Statistic of 226 ERLT-Performance

TL	TGbar	t	P-value
8.2370	12.7626	-3.5032	0.0820

The result of Stout's test of essential unidimensionality showed that the null hypothesis was not rejected (T=-3.5032, p = 0.0820). This was in agreement with the set condition for assessing unidimensionality by (Stout, Habing, Douglas, Kim, Roussos and Zhang, 1996). That null hypothesis is rejected if the test statistic is larger than the predetermined critical value from the normal distribution. This leads to the conclusion that the AT is dimensionally distinct from PT. Otherwise, the test is viewed as essentially one-dimensional.

Which of the IRT- Model for dichotomous information fits the DRA-ERLT?

To answer this question, model - data fit assessment was applied to know which of the model (that is one parameter logistic model and two parameters logistic model) best fit the test items. This was conducted using Jmetrik version 4.1.1. Table 3.7 presents likelihood-based values and goodness of fit Statistics.

Table 3.7 Likelihood-based values and goodness of fit statistics

	Model	
	1PL	**2PL**
-2 Log Likelihood	93192.843	90454.791

Table 3.7 shows the values obtained for -2LogLikelihood (-2LL) for each model that is 1 and 2 parameter logistics model. To establish which model is most appropriate for the data structure, the model with the lowest -2log likelihood value among the models is adjudged to be the most appropriate model-data fit. As the 2-parameter logistic model whose value was the lowest among the models given in Table 3,7. It was concluded that 2-PL model provides significantly better fit when compared to 1PL. Therefore, it was used to establish item parameters of ERLT test items.

3.9 Method of Data Analysis.

Several analyses were performed on the data gathered from the test instrument using Jmetrik software version 4.1.1. Descriptive statistics, empirical reliability, Item Response Theory (IRT) estimation procedures were applied on the data structure where item fit, item and person parameter estimates were established. DIF module of DIFAS software was used in determining the presence of DIF in the groups. Normative data was generated by using T-score and Stannine methods.

3.10 Methodological Challenges

One of the challenges encountered in the course of the study was a kind of restriction from few of the private school owners that were not really convinced that the administration of the test on their pre-school children was just for research purpose. The researcher had to liaise with the local Inspectors of Education that are in charge of the affected schools to educate and convince the affected proprietors/head teachers/caregivers on the exercise. The researcher had to carry out some stimulation activities with the children of the affected schools before the administration of the instrument. Also, all head of schools involved in the study were given feedback on what were observed on their children's response to each section of the test, emphasising on their strong points and giving suggestion on their weak points.

Another challenge was the stress involved in administering the test of 226 items orally on over 700 children one-on-one basis. More research assistants were trained and co-opted in the study after the first three days experience on the field.

Table 3.8: The Research Questions and Methods of Analysis

1	Which of the IRT models for dichotomous data fits the ERLT?	-2log likelihood option of Jmetrik software version 4.1.1
2	What are the estimated parameters of the Draft Early Reading Literacy Test item (DRA-ERLT) using item response theory framework?	Item statistics module of Jmetrik
3	How many of the DRA-ERLT survived using IRT framework?	Number counts
4	Do Draft Early Reading Literacy Test Items (DRA-ERLT) differentiate significantly between; Gender (male and female pre-school children) School location (urban and rural) Type of school (private and public)?	DIF module of DIFAS software using Mantel-Hamzel Method.
5	What are the item and person statistics in the DEV-ERLT using IRT framework?	Mean score, z-score and T-score
6	What are the estimates of reliability for sub-sets Early Reading Literacy Achievement Test?	Empirical Reliability - an output of IRT frame work.
7	What are the range of difficulty levels of identification (of alphabets, objects, animals, birds, human beings and part of the body), recognition of signs and symbols, and colours identification, reading fluency and picture reading subsets in the DEV-ERLT achievement test?	Descriptive statistics
8	Are there normative data developed to facilitate the interpretation of the Early Reading Literacy Test (ERLT) scores with respect to age, gender, school type and school location?	Stannine and T-score

CHAPTER FOUR
RESULTS AND DISCUSSIONS

This chapter presents the results of data analyses and discussion of the findings;

4.1 Research Question One: Which of the IRT- Model for dichotomous data fits the DRA-ERLT?

To answer this question, model-data fit assessment was established to know which of the model (that is one parameter logistic model and two parameters logistic model) best fit the test items. This was conducted using Jmetrik version 4.1.1. Table 4.1 presents likelihood-based values and goodness of fit statistics.

Table 4.1: Likelihood-based values and goodness of fit statistics

	Model	
	1PL	**2PL**
-2 Log Likelihood	93192.843	90454.791

Table 4.1 shows the values obtained for -2LogLikelihood (-2LL) for each model that is 1 and 2 parameter logistics model. To establish which model is most appropriate for the data structure, the model with the lowest -2log likelihood value among the models is adjudged to be the most appropriate model-data fit. As the 2-parameter logistic model whose value was the lowest among the models given in Table 4.1, it was concluded that 2-PL model provides significantly better fit when compared to 1PL. Therefore, it was used to establish item parameters of ERLT test items.

Discussions

The result is in accordance with Thorpe and Favia (2012) that higher values indicate a poorer fit of the model to the data and comparing the values from different models can indicate which model represents better fit. The result agrees with Essen(2015) that examined which of the model fit best the data of scores in 50 items in UTME Mathematics that was scored dichotomously, using BILOG MG 3.0 software. The result revealed that no item fitted the 1 parameter model and the 3 parameter model while 26 items fitted 2 parameter model. The conclusion was that the 2 Parameter logistic model was the best fit for the data.

The result also corroborates Metibemu (2016) research work where 100 items PAT were subjected to analysis using BILOG MG. The items best fitted 2-PL model. On the other hand, the result contradicts Wiberg (2004) who examined the 3 Items Response Theory with the scores of the examinees to choose which of the IRT model fit the data best. In the result of the analysis, 3- parameter logistic model came out the best.

4.2 Research Question Two: What are the estimated parameters of the Draft Early Reading Literacy Test item (DRA-ERLT) using Item Response Theory framework?

The test items contained in the DRA-ERLT which were subjected to item calibration using Jmetrik software. As a result of the model fit assessment conducted, the items fitted 2-PL model. The outputs were from the 2-PL model representing IRT procedure of estimation. Thus, Table 4.2 presents statistics for the IRT item parameters where *b* represents the difficulty parameter, *a* represents the discrimination parameter.

Table 4.2: The item parameter estimates in the DRA-ERLT using IRT framework

ITEM NUMBER	*IRT*	
	a	***b***
ERLT 1	0.61	-2.20
ERLT 2	1.17	-1.23
ERLT 3	1.53	-2.50
ERLT 4	2.20	-1.84
ERLT 5	2.96	-1.62
ERLT 6	2.85	-1.42
ERLT 7	3.25	-1.07
ERLT 8	3.34	-1.00
ERLT 9	4.30	-0.95
ERLT 10	3.12	-0.88
x x x		
ERLT 217	1.89	-3.40
ERLT 218	1.90	-1.05
ERLT 219	0.88	-5.69
ERLT 220	1.17	-3.84
ERLT 221	1.03	0.46
ERLT 222	2.81	0.73
ERLT 223	2.49	-4.51
ERLT 224	3.05	0.53
ERLT 225	1.35	0.10
ERLT 226	1.12	0.77

xxx Abridged version of parameter estimates of ERLT 11-216. For full table see Appendix IV

Table 4.2 shows that columns one and two gave the discrimination (a) and difficulty (b) parameters of Item Response Theory model obtained from Jmetrik software output. It was observed from the result that the IRT framework gave estimates of all the item parameters of the 226 DRA-ELRT items subjected to calibration process. This implies that test items fitted 2- parameter logistic model. However, for item discrimination (a), the values ranges between 0.12 (item 112- Identification of basket) and 6.36 (Item 21 - Identification of letter U)

Discussions

From the result in column 1 of Table 4.2, it was deduced that five items were below 0.34, seven items were within the range of 0.35 and 0.64, 31 items were between 0.65-1.34, while 20 items were within 1.35 – 1.69. These are explained in Tables 4.2(i), 4.2(ii), 4.2(iii) and 4.2(iv) respectively.

Table 4.2(i): Items of very low discrimination power

Discrimination range	No. of Items	Items	Interpretation
Below 0.34	5	90, 91, 113, 121 and 166	Very low discrimination power between pre-school children of high and low ability

Table 4.2(ii): Items of low discrimination power

Discrimination range	No. of Items	Items	Interpretation
0.35 – 0.64	7	76, 112, 159, 196, 205, 209 and 210	Low discrimination power could not really differentiate between the children that are of high and low ability.

Table 4.2(iii): Items of moderate discrimination power

Discrimination range	No. of Items	Items	Interpretation
0.65 – 1.34	31	12, 32, 43, 53, 54, 101, 103, 116, 142, 150, 151, 153, 162, 164, 168, 170, 189, 198, 199, 200, 201, 202, 203, 204, 206, 213, 215, 216, 220, 221 and 226	Moderate discrimination power.

Table 4.2(iv): Items of high discrimination power

Discrimination range	No. of Items	Items	Interpretation
1.35 – 1.69	20	3, 27,58, 77, 83, 84, 101, 107, 114, 122, 133, 134, 141, 154, 163, 181, 184, 198, 207 and 208	High discrimination power showed the clear differences in pre-school children ability in Early Reading Literacy Skills

It was observed that many of the items that exhibited higher discrimination index fall majorly within the subset of identification of alphabets, some on identification of objects and one item on reading fluency - "It is a window". It is important to mention that this same item on reading fluency was adjudged the most difficult from the result of the difficulty level of the reading fluency subset of DEV-ERLT after its administration on 776 pre-school children. This agrees with the findings that the item highly discriminates between the preschoolers of high and low abilities. The reason for having the reading of the sentence "It is a window" as the most difficult item might be because window is a six letter word that children are not really familiar with the spelling. It might also be due to the fact that letter "w" is one of the letters many examinees were having difficulty in its identification. The item also has the highest discrimination index probably because only few of the examinees that are of high ability got it right, it may likely be that they are opportune to learn this from any source outside school. The implication of this is that pre-school teacher should pay more attention to the exposition of preschoolers to the identification, pronunciation of letter "w" and some objects that carry "w" in their spellings. They should expose children to different types of windows for better performance. The remaining items fall within 1.76 (item 30 - Identification of letter D) and 6.36 (item 21 - Identification of letter U). This is interpreted as having a very high discrimination index.

The range for discrimination is between 0 and 2. Baker (2001) submitted that items with discrimination below 0.34 are considered to have very low discrimination, items between 0.35 and 0.64 are low, 0.65 and 1.34 are considered moderate, 1.35 and 1.69 are high while 1.70 and above are very high. Also, for item difficulty parameter (b), it ranges between -7.92 and 0.85. The items with lower b values are considered to be easier than items with higher b values. However, Baker (2001) argues that the typical range for difficulty is between -3 and +3. Thus, items in the ERLT that have

values greater than 3 can be considered as extremely difficult while items with values less than -3 can be regarded as very cheap.

Thus, item 195 has 0.85 difficulty level while item 119 has the difficulty level of -7.92. This indicates that item 195 (The reading fluency of "She is a girl") is the most difficult, probably, because at this level the children are not really familiar with the pronoun "site" they are most familiar with "IT". This corroborates the observation on the field of the study. On the other hand, item 119 (identification of a leaf) is adjudged the easiest item in the result. Children see leaf all around them, touch and play with it, this might have enhance their ability to identify it in the test. It was found that larger percentage of the items in DRA-ERLT fall within the acceptable range of difficulty. This submission laid credence to the work of (Fakayode, 2017) that all the Mathematics items in the June test falls within this range of difficulty while only some items in the November test falls within the range. The implication of the result is that teachers handling preschoolers should ensure they expose them to other forms of pronoun like "SHE" ," HE" and "THEY" . Pre - school teachers need not to dwell much on items that appear too easy but should devote that mere attention and time to those difficult.

4.3 Research Question Three: How many of the items of the DRA-ERLT survived using IRT framework?

Table 4.3: The item parameter estimates and survived Items

ITEM NUMBER	*IRT*		
	a	***b***	***Remarks***
ERLT 1	0.61	-2.20	Selected
ERLT 2	1.17	-1.23	Selected
ERLT 3	1.53	-2.50	Selected
ERLT 4	2.20	-1.84	Selected
ERLT 5	2.96	-1.62	Selected
ERLT 6	2.85	-1.42	Selected
ERLT 7	3.25	-1.07	Selected
ERLT 8	3.34	-1.00	Selected
ERLT 9	4.30	-0.95	Selected
ERLT 10	3.12	-0.88	Selected
x			
x			
x			
ERLT 216	1.13	0.38	Selected
ERLT 217	1.89	-3.40	Rejected
ERLT 218	1.90	-1.05	Selected
ERLT 219	0.88	-5.69	Rejected
ERLT 220	1.17	-3.84	Rejected
ERLT 221	1.03	0.46	Selected
ERLT 222	2.81	0.73	Selected
ERLT 223	2.49	-4.51	Rejected
ERLT 224	3.05	0.53	Selected
ERLT 225	1.35	0.10	Selected
ERLT 226	1.12	0.77	Selected

xxx Abridged version of parameter estimates of ERLT 11-216. For full table see Appendix IV

It can be observed from Table 4.3 that column 4 gave remarks on DRA-ERLT items that fall within an acceptable range of difficulty and those that did not meet the criteria. Careful estimate of these items depicts that there were sixty-five (65) items that fell out of the range. These are: item 22, 24, 28, 30, 34, 35, 36, 43, 44, 45, 60, 79,

80, 82, 83, 84, 88, 89, 90, 92, 96, 97, 98, 101,110, 116, 119, 129, 140, 141, 145, 147, 148, 157, 158, 159, 162, 163,164,165, 166, 169, 170, 174, 176, 177, 178, 179, 181, 183, 184, 186,189, 190, 191, 194, 196, 203, 204, 205, 206, 207, 217, 219,220 and 223. While those items that met the difficulty range of -3 to +3 were selected. Figure 4.1 shows the bar chart of selected and deleted items.

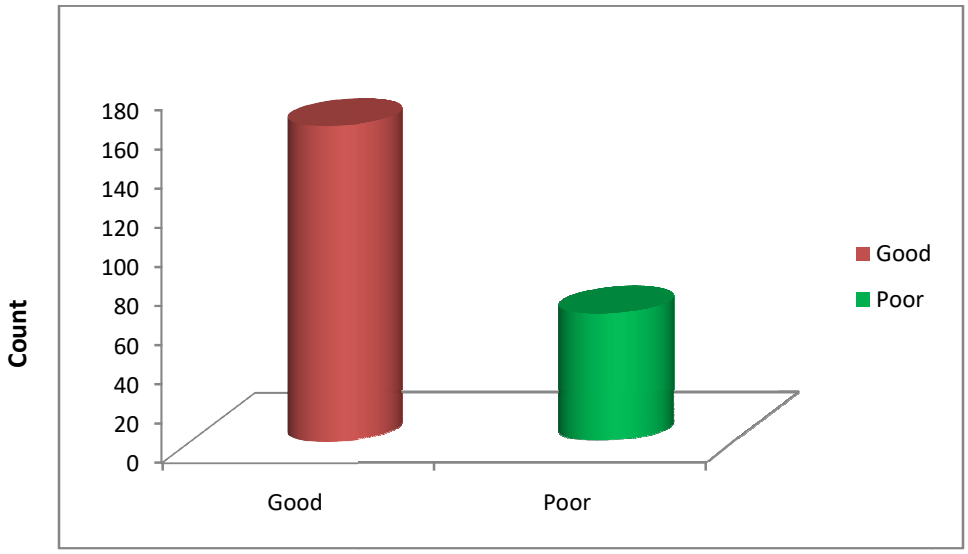

Figure 4.1: Number of selected and deleted Items

Discussions

These are the list of some items that were adjudged poor: Q, W, S, h, b, q, y, d, i and f. Identification of door, clock, snake, pineapple, monkey and frog. It is necessary to mention that all items on identification of shapes did not survive and this agrees with the observation on the field that only 5% of the examinees could identify circle shape among the four items that tested shape identification. The reason for the poor performance of examinees on the aspect of identification of shape is that probably 90% of the pre-schoolers have not been exposed to this aspect of identification in their centres. The teachers of the sampled pre-schools also confirmed that they have not exposed the children to this aspect despite the fact that the administration of the test took place towards the end of third term .The implication is that the items of the subset on the identification of shapes are too difficult for the examinees level as at the time of this study and that the teachers should ensure they stimulate the pre-school children to learn the identification of common shapes in their

environment through different methods because it is required of them at this level as stipulated in the curriculum.

More importantly, it is noteworthy to state that item numbers 22 and 24 were statistically adjudged as rejected, (that is, alphabets V and X). The researcher did not delete the two items because of their roles in the sub-set. English language alphabets cannot be completely presented without letter V and X. The purpose of the test sub-set was to assess the pre-school children's ability in the identification of letters in English language which are 26 in number, removal of letter V and X will jeorpardise the purpose of the test. This is in line with Adegoke (2013) submission that item selection under the IRT is based on the intended purpose of the test.

Letter V and X did not survive the IRT analysis probably because pre-school children are having difficulty in the identification of these letters. Many of them pronounce 'X' as 'S', 'V' as 'F' probably due to cultural linguistic interference .While many due to rote learning omitted 'V', while reciting the English language letters and eventually were not able to identify it while the test was administered on them.

4.4 Research Question Four: Do Draft Early Reading Literacy Test Items (DRA- ERLT) differentiate significantly between;

a) Gender (male and female pre-school children)

b) School type (urban and rural)

c) School location (private and public)?

To establish items that function differentially between male (reference group) and female (focal group) examinees, the DIF option of DIFAS (Differential Item Functioning Analysis System) software version 5.0 was conducted. The analysis showed that value of Standardised Mantel-Haenszel Log-Odds Ratio (LOR Z) greater than ∓2.0 is considered as evidence of presence of DIF and if otherwise NO DIF. Also, zero value indicates an absence of DIF, a negative value indicates that the item favours the reference group, while positive value indicates that items favours the focal group. The results of the analysis of DIF for the DRA-ERLT items with respect to demographic information such as gender, school location and school type were presented in Tables 4.4, 4.4.1 and 4.4.2 respectively.

Table 4.4: The analysis of DIF with respect to Gender

Item number	M-H CHI	M-H LOR	LOR SE	LOR Z	BD	ETS
Item 1	0.00	0.00	0.00	0.00	0.00	NO DIF
Item 2	0.01	-0.60	0.99	-0.60	1.54	NO DIF
Item 3	0.00	-0.35	0.81	-0.43	0.34	NO DIF
Item 4	0.05	0.39	0.69	0.56	0.06	NO DIF
Item 5	0.03	0.08	0.66	0.12	1.41	NO DIF
Item 6	0.82	-0.57	0.51	-1.12	2.61	NO DIF
Item 7	0.00	0.09	0.46	0.19	0.49	NO DIF
Item 8	0.68	-0.44	0.42	-1.04	0.38	NO DIF
Item 9	0.82	-0.54	0.47	-1.14	2.84	NO DIF
Item 10	0.02	0.03	0.44	0.07	0.43	NO DIF
x						
x						
x						
Item 216	2.21	-0.38	0.24	-1.58	3.47	NO DIF
Item 217	0.02	0.01	0.33	0.04	0.01	NO DIF
Item 218	0.28	0.28	0.38	0.73	0.17	NO DIF
Item 219	0.08	-0.11	0.26	-0.42	0.02	NO DIF
Item 220	0.02	0.01	0.30	0.02	0.02	NO DIF
Item 221	1.21	0.32	0.26	1.24	0.01	NO DIF
Item 222	0.33	0.45	0.54	0.82	0.01	NO DIF
Item 223	0.55	-0.53	0.51	-1.04	0.48	NO DIF
Item 224	0.01	0.17	0.51	0.33	1.96	NO DIF
Item 225	0.00	0.02	0.27	0.08	0.01	NO DIF
Item 226	0.31	-0.25	0.34	-0.75	0.54	NO DIF

xxx Abridged version of DIF with respect to gender (ERLT 11-215). For full table see Appendix IV

Figure 4.2 shows the line graph of items that indicates DIF in favour of males and females:

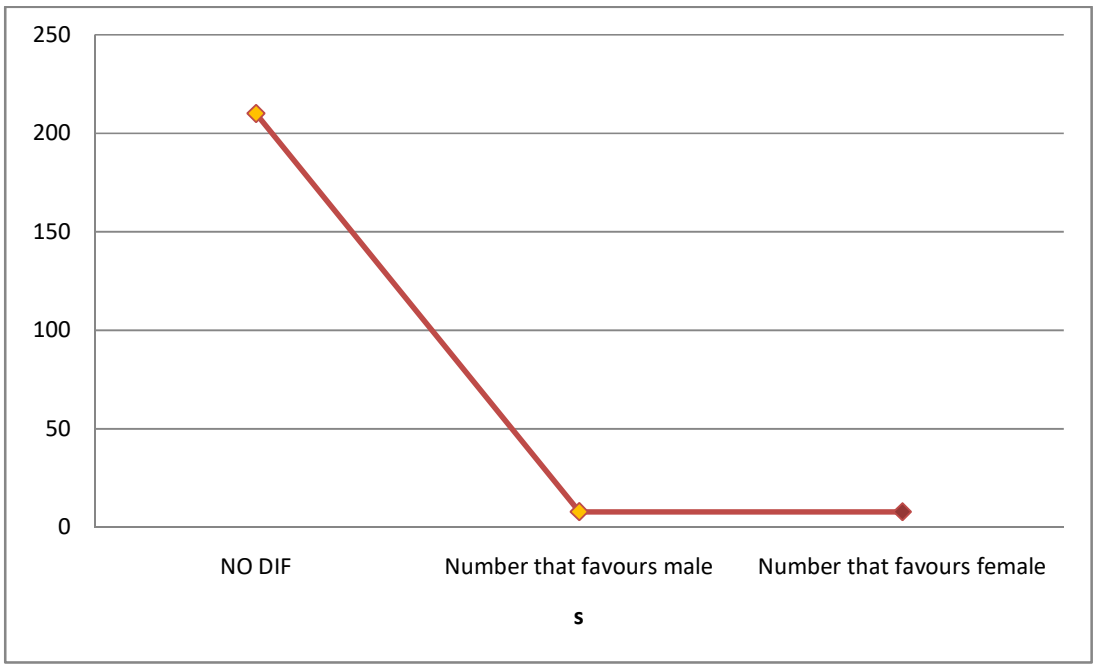

Figure 4.2: Number of items that favoured males and females

Table 4.4 depicts the DIF statistics on examinees item performance with respect to gender. Column 2 of Table 4.4 is the Mantel Haenszel Chi-square (MH CHI). This is distributed as chi-square with one degree of freedom (Holland and Thayer, 1988). While column 3 are Mantel Haenszel Common Loggs - odds Ratio (MH LOR) , it is asymptotically normally distributed where negative values means the reference groups is favoured in the DIF, while positive values indicate presence of DIF in favour of the groups.

Column 4 is the standard error of the Mantel-Haenszel Common LOg - Odds Ratio (LOR SE). Non symmetric estimator presented by Robins, Breslow and Greenland (1988) was computed as the standard error. Column 5 is the Standardised Mantel-Haenszel Logg- Odds Ratio(LOR Z), it is the division of Mantel-Haenszel Log-odds ratio by the estimated standard error (Camilli and Shepard, 1994). Column 6 is Breslow - Day Chi-square (BD). The BD chi-square of trend in odds ratio heterogeneity is distributed as chi-square with one degree of freedom. Column 7 shows the ETS, this is used to bring out those items that exhibit DIF.

However, Table 4.4 revealed that 16 items have standardised Mantel-Haenszel Log-Odds Ratio (LOR Z) values greater than ∓ 2.0. These items are; 24, 39, 44, 70, 84, 107, 118, 125, 132, 133, 142, 146, 149, 172, 173. Out of 226 items only 16 items show difference in gender(8 favoured males, 8 favoured females) that is 31.0% of all the DRA-ERLT items.

Discussions

It can be deduced that the test had not favoured males against females' examinees. This supports Igbokwe (2004) who found that there were no significant differences in Mathematics Performance between boys and girls when the researchers developed item bank in Mathematics for NECO common entrance test. Also, the finding corroborates Falayajo, Mokoju, Okebukola, Onugah and Olubodun (1997) results which assessed the level of competency of primary four pupils in numeracy, literacy and life skills. The sample was twenty-three thousand and forty pupils, the results revealed there was a very little difference in the mean scores of boys- 24.8 percent and girls - 25.8 percent.

The result on the other hand, contradicts Bagneto and Neisworth (2010) report that girls performed better than boys in literacy test. Bagneto and Neisworth (2010) attempted to give some explanations to this girl's superiority in language development that it is partly caused by their closeness to their mother even from birth where they enjoyed emotional satisfaction and stability in an ideal situation and opportunity to listen to their mother speak and repeat after them. Also, the study contradicts the work of Ijaiya (2007) who found that females performed significantly better on word fluency test. The result of Bagneto and Neisworth (2010) study may likely due to the fact that boy fetal brain develops an asymmetry of the two hemispheres with a growth and size of both left and right hemispheres remaining equal while girl fetal brain develops an asymmetry in which more rapid and greater neural development of the left hemisphere temporal lobe (that is, language areas).

It is important to operationalise the DIF result in the study, the items that favoured one gender than other. For example, identification of a bag appeared to favour female pre-school children because the object is gender specific, females are more attached to the use of bags than males. Carrying of bags is seen as part of females' dressing while males rarely carry bags, many of them are only used to their

school bags. In the same vein, identification of a tree favoured males examinees, this could be as a result of the cultural belief that only males are capable of climbing, playing around and on the trees. Most of the time, are opportune to see more trees of different sizes and shapes of trees, girls are not always allowed to participate in activities that will require climbing trees.

It is noteworthy to state that those items that exhibited DIF statistically were not expunged from the 226 DRA-ERLT on the basis of bias during the development process of the ERLT. This laid credence on the submission of Wiberg (2004); Metibemu (2016) that before an item is removed from a test based on DIF, it must be analysed by subject matter expert to know what actually went wrong with the item. The implication of this result is that teachers handling pre-schoolers should ensure the children are exposed to different examples that are gender balanced and all the children should participate in all the activities of learning irrespective of their gender. The teachers should also pay more attention to a particular gender that an item seems not to favour in the study.

Table 4.4.1: Analysis of DIF with respect to School Type

Item number	MH CHI	MH LOR	LOR SE	LOR Z	BD	ETS	Remarks
Item 1	0.00	0.00	0.00	0.00	0.00	NO DIF	
Item 2	0.37	-0.05	1.05	-0.05	3.72	NO DIF	
Item 3	3.05	0.00	0.00	0.00	0.00	NO DIF	
Item 4	4.13	-1.59	0.76	-2.11	0.39	DIF	Favour private examinees
Item 5	1.88	-0.92	0.59	-1.55	0.27	NO DIF	
Item 6	0.98	-0.54	0.47	-1.16	1.38	NO DIF	
Item 7	0.82	-0.47	0.43	-1.09	1.19	NO DIF	
Item 8	2.69	-0.65	0.39	-1.65	0.01	NO DIF	
Item 9	2.44	-0.85	0.50	-1.72	11.90	NO DIF	
Item 10	2.11	-0.89	0.54	-1.65	10.07	NO DIF	
x							
x							
x							
Item 216	1.93	0.43	0.30	1.43	0.46	NO DIF	
Item 217	2.39	-0.59	0.38	-1.55	5.11	NO DIF	
Item 218	0.00	0.06	0.39	0.15	0.00	NO DIF	
Item 219	39.12	1.55	0.29	5.38	24.88	DIF	Favour public examinees
Item 220	7.77	-1.10	0.44	-2.54	8.56	DIF	Favour private examinees
Item 221	10.39	0.92	0.29	3.22	5.49	DIF	Favour public examinees
Item 222	4.78	-1.21	0.59	-2.05	10.36	DIF	Favour private examinees
Item 223	5.17	-1.41	0.61	-2.31	3.13	DIF	Favour private examinees
Item 224	6.06	-1.03	0.57	-1.80	43.51	NO DIF	
Item 225	0.04	-0.11	0.31	-0.35	1.43	NO DIF	
Item 226	0.16	-0.20	0.38	-0.53	0.92	NO DIF	

xxx Abridged version of DIF with respect to school type (ERLT 11-215). For full table see Appendix IV

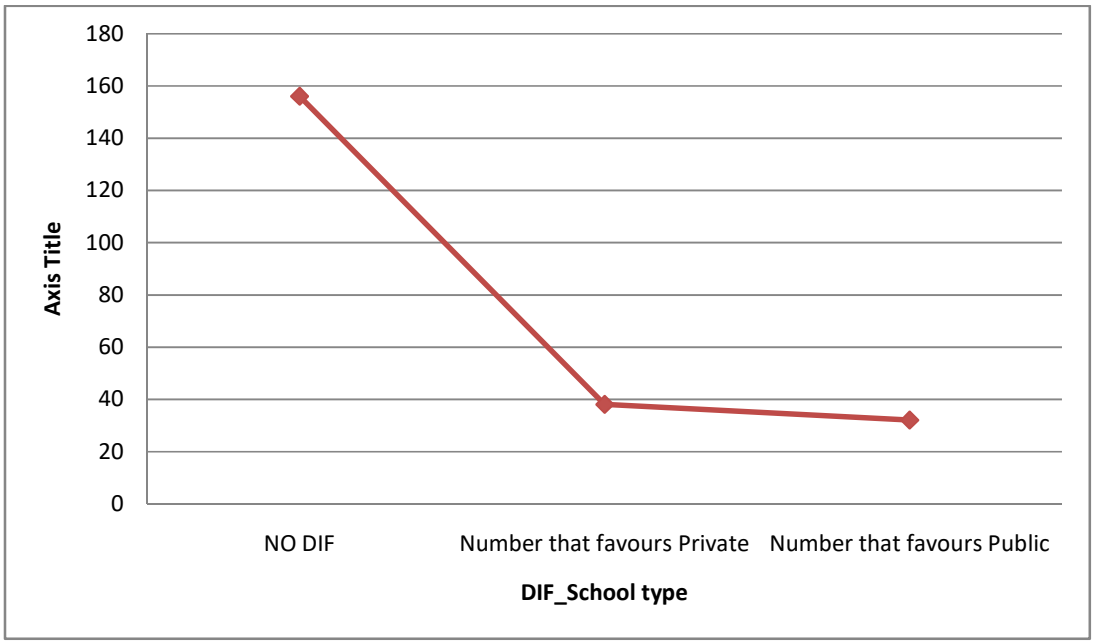

Figure 4.3: Numbers of items that favoured public and private schools

Table 4.4.1 shows the DIF statistics on examinees item performance with respect to school type. Table 4.4.1 indicated that 70 items have standardised Mantel-Haenszel Log-Odds Ratio (LOR Z) values greater than ∓ 2.0.

These items are; 4, 23, 24, 31, 52, 83, 85, 90, 105, 106, 107, 108, 113, 115, 116, 119, 122, 123, 126, 133, 134, 137, 139, 140, 141, 142, 145, 146, 147, 148, 149, 155, 156, 157, 158, 159, 160, 161, 162, 164, 166, 167, 171, 174, 176, 178, 180, 181, 186, 187, 191, 193, 194, 197, 198, 199, 200, 201, 202, 209, 210, 211, 212, 213, 214, 219, 220, 221, 222 and 223.

Discussions

Thirty- seven(37) out of 70 items favoured private schools that is, 52.9% of the items that exhibited DIF while 33 items favoured public schools, that is, 47.1% of the items that exhibited DIF between examinees from private and public pre-schools. Researcher operationalised, the items that were found to exhibit DIF among the examinees in private and public centres. For example, identification of tyre favoured public school examinees. This might be because some male children in the public schools do play with used tyres which might give them better opportunity of identifying tyre than their mates in the private schools. On the other hand, identification of a jug favoured the examinees in the private schools, the reason might

be that many pre-school children in public schools are from family backgrounds that may not give them opportunity of having or using jug at home and so its identification may be difficult unlike many children that attend private schools. These items should not be discarded but it implies that the teachers should ensure these pre-school children are exposed to as many objects as possible that could be found at home and in the environment irrespective of their families economic status, academic background and location.

More importantly, the items that were statistically found to exhibit DIF were not removed from the DRA-ERLT on the account of school type bias during the development process of the ERLT. This finding supports the submission of Nworgu (2011) that the presence of DIF in an item does not necessarily mean that the item is biased. It can also be deduced from the result going by the number count that the examinees from private pre-schools performed better in many of the ERLT items than those from public pre- schools.

Table 4.4.2: Analysis of DIF with respect to school location

Item number	MH CHI	MH LOR	LOR SE	LOR Z	BD	ETS	Remarks
Item 1	0.00	0.00	0.00	0.00	0.00	NO DIF	
Item 2	2.44	0.00	0.00	0.00	0.00	NO DIF	
Item 3	7.98	0.00	0.00	0.00	0.00	NO DIF	
Item 4	2.85	-1.21	0.65	-1.86	0.41	NO DIF	
Item 5	0.85	-0.74	0.64	-1.15	0.26	NO DIF	
Item 6	0.98	-0.81	0.63	-1.29	0.00	NO DIF	
Item 7	9.94	-1.66	0.56	-2.95	4.25	DIF	Favour urban examinees
Item 8	11.47	-1.77	0.57	-3.10	1.76	DIF	Favour urban examinees
Item 9	3.85	-1.12	0.56	-2.02	3.66	DIF	Favour urban examinees
Item 10	4.86	-1.59	0.71	-2.24	0.53	DIF	Favour urban examinees
x							
x							
x							
Item 216	12.35	1.06	0.33	3.18	13.83	DIF	Favour rural examinees
Item 217	3.64	0.76	0.43	1.77	5.81	NO DIF	
Item 218	0.02	-0.07	0.51	-0.13	0.23	NO DIF	
Item 219	0.00	0.03	0.28	0.13	1.51	NO DIF	
Item 220	0.05	0.29	0.57	0.50	0.00	NO DIF	
Item 221	18.53	-1.57	0.39	-4.08	0.08	DIF	Favour urban examinees
Item 222	0.19	0.43	1.60	0.27	0.01	NO DIF	
Item 223	0.08	0.07	0.77	0.09	0.12	NO DIF	
Item 224	0.04	-0.57	0.93	-0.62	1.55	NO DIF	
Item 225	4.24	-0.77	0.38	-2.04	0.67	DIF	Favour urban examinees
Item 226	0.60	0.57	0.58	0.98	0.22	NO DIF	

xxx Abridged version of DIF with respect to school location (ERLT 11-215). For full table see Appendix IV

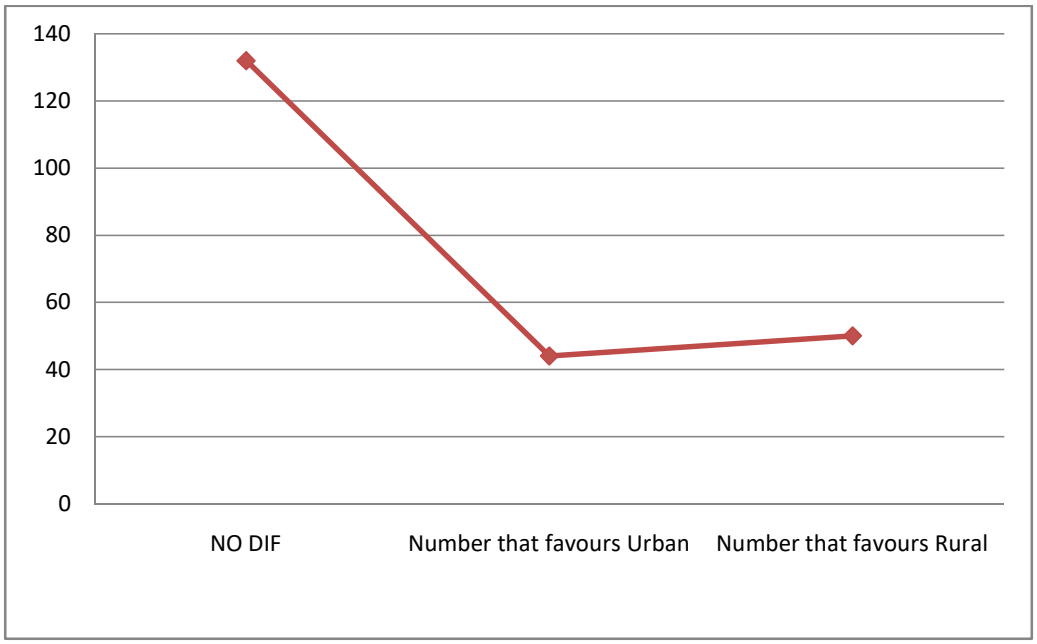

Figure 4.4 Number of items that favoured pre-schools in the rural and urban

Table 4.4.2 shows the DIF statistics on examinees item performance with respect to school location. However, table 4.4.2 indicated that 94 items have standardised Mantel-Haenszel Log-Odds Ratio (LOR Z) values greater than ∓ 2.0. These items are; 7, 8, 9, 10, 11, 14, 20, 23, 25, 29, 31, 35, 36, 37, 40, 41, 42, 44, 46, 47, 49, 51, 52, 59, 60, 61, 62, 69, 70, 71, 76, 81, 83, 84, 91, 92, 95, 96, 98, 99, 100, 102, 105, 110, 111, 112, 113, 117, 118, 120, 121, 125, 130, 132, 133, 134, 137, 139, 142, 144, 146, 150, 152, 159, 162, 164, 166, 169, 170, 171, 175, 176, 178, 179, 180, 182, 186, 190, 191, 196, 203, 204, 205, 206, 207, 208, 209, 209, 210, 213, 214, 215, 216, 221 and 225. that is 42% of the 226 items exhibited DIF in school type.

Discussions

The results also showed that 44 out of 94 items that is 46.8% of all the items that exhibited DIF favoured urban examinees, while 50 items that is, 53.2% of the items that exhibited DIF favoured rural examinees. It can be deduced that more items favoured rural examinees. Item on the identification of an egg favoured urban examinees. This might be because many pre-school children in the urban setting see egg often around them and take egg as part of their meals unlike many examinees in the rural who rarely take egg as part of their meal. On the other hand, identification of a ladder favoured rural examinees, this might be because examinees in the urban may

not have the opportunity of seeing a ladder before unlike many children in the rural areas who see people climbing ladder probably to adjust any damage on the roof of their houses or for some other purposes.

It is important to state that none of the items was removed from the 226-item DRA-ERLT on the basis of location bias during the development process of the ERLT. This is premised on the submission of Metibemu (2016) that before an item is removed from a test based on DIF it must be subjected to subject matter expert review to know what actually went wrong with the item. The implication of the result of this study is that pre-school teachers should ensure that they expose children to all these objects that could be found in the rural and urban settings irrespective of the pre-schoolers location. Teachers should try as much as possible to balance the examples of objects, fruits and others they use in the centres in stimulating the pre-schoolers to learn not minding the extra efforts it takes.

4.5 Research Question Five: What are the item and person statistics in the DEV- ERLT using IRT framework?

Table 4.5: Item parameters in the DEV-ERLT using IRT framework

ITEM	IRT	
	a	***b***
ERLT0001	0.936	-4.472
ERLT0002	1.177	-3.133
ERLT0003	1.574	-2.494
ERLT0004	2.175	-1.886
ERLT0005	2.832	-1.691
x		
x		
x		
ERLT0157	0.972	-0.853
ERLT0158	2.973	0.455
ERLT0159	3.277	0.527
ERLT0160	1.311	0.123
ERLT0161	1.164	0.764

xxx Abridged version of Item and person statistics in DEV-ERLT 0006 - 0156. For full table see Appendix IV

Table 4.5 depicts the discrimination (a) and difficulty (b) parameters of Item Response Theory model. It was observed from the result that the IRT framework gave estimates of all the item parameters in the 161 DEV-ELRT items subjected to calibration process. This implies that test items fitted 2- parameter logistic model. However, for item discrimination (a), the values ranges between 0.13 and 6.35 with (M=2.57; SD= 1.34). The range for discrimination is between 0 and 2. Baker (2001) submitted that items with discrimination below 0.30 are considered to have low discrimination, items between 0.30 and 1.34 are moderate, 1.35 and 1.69 are considered high, while 1.70 above are very high. Also, for item difficulty parameter (b) ranges between -6.90 and 0.95 with (M=-0.80; SD= 1.0635). The items with lower b values are considered to be easier than items with higher b values. However, the typical range for difficulty is between -3 and +3. Thus, items in the ERLT test that

have values greater than 3 can be considered as extremely difficult, while items with values less than -3 can be regarded as very simple. From the result, it can be said that ERLT items discriminate better among examinees of low and high ability.

However, ability scores in the DEV-ERLT test under IRT was examined. Tables 4.5.1 and 4.5.2 presents the ability scores and descriptive statistics of ERLT test under IRT framework

Table 4.5.1: Statistics of ability scores in the DEV-ERLT using IRT framework

Examinees	Ability score
1	-1.333
2	0.445
3	-1.333
4	-0.444
5	0.444
6	-1.333
7	0.444
8	0.444
9	-0.441
10	-0.410
x	
x	
x	
767	-1.333
768	-1.333
769	0.423
770	-1.203
771	0.452
772	-1.352
773	-1.335
774	-0.444
775	-2.218
776	-0.444

xxx Abridged version of ERLT 11-766. For full table see Appendix IV

Table 4.5.2: Descriptive statistics of ability scores in the DEV- ERLT

Statistics	N	IRT	
		ERLT- TEST	Tscore
Min.	776	-3.02	19.80
Max.	776	2.17	71.70
Mean	776	0.11	51.10

Discussions

From Table 4.5.1 and 4.5.2, it could be deduced that 776 examinees ability scores ranged between -3.02 and 2.17. The examinees' ability estimate (in z-score) under IRT was transformed to T-scores. Then the overall mean test score was obtained. It was observed from Table 4.4.5 that the highest score was 71.70, the least was 19.80. While the mean score was 51.10, It can be said that many of the examinees performed above average. This also implies that on the average, the test items were neither too easy nor too difficult for examinees within the norm.

On the average, the preschoolers examined performed a little above average in the Early Reading Literacy Test probably, because the items in the instrument were in accordance to the approved curriculum for the level of the examinees and it was observed that the capacity of few of the teachers handling them have been built on the better ways of stimulating the preschool children to learn. Also, there was a little improvement in the using of play materials, picture books that were locally made and improvised by some teachers in stimulating the preschoolers to develop some of the pre-skills necessary in improving their reading literacy skills. Large percentage of children involved in the study were happy to be in the preschools and learnt better through play. Some of the reasons highlighted, despite the many other factors that may have negative implication on the preschoolers learning rate might have contributed to the result that manyt examinees performed a little above average in this study. This corroborates the result of National Reading Panel (2009), reading test administered on 4th grade children in the United States showed that the mean score was 52.10 which means large percentage of the examinees performed a little above average.

4.6 Research Question Six: What are the estimates of reliability for sub-sets Early Reading Literacy Achievement Test?

Table 4.6: Reliability coefficients of sub - sets of Early Reading Literacy Test

S/N	Sub-sets	Reliability Coefficient
1	Identification of letters	0.99
2	Identification of objects, animals, parts of human body and the rest	0.98
3	Identification of colours	0.62
4	Recognition of signs and symbols	0.93
5	Reading fluency	0.98
6	Rhymes	0.91
7	Picture Reading	0.96

Table 4.6 shows that the reliability indices of the sub-sets range from estimate reliability of IRT = 0.62 to 0.99, while the identification of letters has the highest reliability coefficient value, identification of color has the least coefficient value.

Discussions

In general, all the test items in the DEV-ERLT indicate a high degree of internal consistency. Likely reason for the identification of letters having the highest reliability coefficient value is that it is the most important and first sub-skills necessary in developing reading literacy skills, therefore, a pre-schooler among other things is first exposed to the identification of letters through different means; songs, rhymes, pictures and the rest. English language letters are the same irrespective of location or time, this means any child that is able to identify these letters might not likely get it wrong as long as the same English alphabets are presented to the child. This corroborates assertion of Chall (1993) that learning to read starts with learning the code, (that is, letters) and that stage one in learning to read is pre- reading which is learning the code(letters).

4.7 Research Question Seven: What are the range of difficulty levels of identification (of alphabets, objects, animals, birds, human beings and part of the body), recognition of signs and symbols, and colours identification, reading fluency and picture reading subsets in the DEV-ERLT achievement test?

Table 4.7: Difficulty level of identification of Alphabets

ITEM	IRT	
	A	b
IA1	0.936	-4.472
IA2	1.177	-3.133
IA3	1.547	-2.494
IA4	2.175	-1.886
IA5	2.832	-1.691
IA6	2.553	-1.505
IA7	3.225	1.124
IA8	3.183	-1.041
IA9	4.058	-0.904
IA10	2.894	-0.904
IA11	4.940	-0.938
IA12	4.678	-0.849
IA13	4.4732	-0.813
IA14	5.477	-0.808
IA15	4.944	-0.865
IA16	4.647	-0.723
IA17	4.502	-0.578
IA18	4.984	-0.622
IA19	4.421	-0.685
IA20	4.828	-0.628
IA21	6.325	-0.578
IA22	0.668	-0.544
IA23	2.022	-0.862
IA24	3.065	-0.895
IA25	1.478	-2.409
IA26	3.912	-0.480

Discussions

Table 4.7 shows the difficulty level of identification of letters which value ranged from -4.47 to 0.48. This indicates that the items are too cheap. The likely reason for this is that identification of alphabets is the first content a child is exposed to as soon as the child can talk even sometimes from home before enrolling for pre-school. Observation on the field also revealed that there was no examinee that could not read and identify some of the alphabets. The number count of examinees that got the items on alphabet identification right ranged between 543 and 774 out of 776 examinees. This implies teachers need not too spend much time in this aspect but should lay emphasis on the few that letters the children are having difficulty in its identification.

Table 4.7.1: The Difficulty Levels of Identification of Objects, Animals, Fruits and Parts of Human Body

ITEM	IRT	
	a	b
IDEN01	1.526	-0.206
IDEN02	3.01	-0.593
IDEN03	2.605	-0.462
IDEN04	2.354	-0.374
IDEN05	0.52	-4.861
x		
x		
x		
IDEN41	1.494	-1.562
IDEN42	2.179	-0.826
IDEN43	2.062	0.227
IDEN44	2.039	-0.777
IDEN45	1.743	-0.955

xxx Abridged version of ERLT 06 - 40. For full table see Appendix IV

Table 4.7.1 shows difficulty levels of all the items in the identification of objects, fruits animals, birds and parts of human body. These were calibrated using 2-PL. The item difficulty parameter (b) ranges between -6.90 (item 12 - door) and 3.99 (item 8 - orange) which implies that item 12 is the easiest and item 8 is the most difficult. The results indicate that identification of door is the simplest of the sub-set,

while identification of orange is the most difficult. The likely reason for having the identification of door as the very easy among others in this subset is that every day pre-school children have contact with the door, they enter their houses and classrooms through the door. They also sing not less than two songs in a day where they mention, point to or knock the door. Therefore, they develop vocabulary through their daily interactions. This supports Kaderavek and Pertimonten (2014) claim that children acquire most of what they know about oral language by interacting, listening and speaking with others, in the process, they build vocabulary(ies) that form the foundation for reading skills.

Table 4.7.2: The Difficulty Levels of Reading Fluency

ITEM	IRT	
	a	b
RF01	1.977	-0.557
RF02	2.125	-0.412
RF03	1.708	-0.135
RF04	1.669	0.043
RF05	3.908	-0.318
RF06	3.241	0.236
RF07	3.514	0.212
RF08	3.862	0.234
RF09	4.120	-0.266
RF10	3.283	0.037

Table 4.7.2 shows difficulty levels of all the items contained in reading fluency of the Developed ERLT which were calibrated using 2-PL. The item difficulty parameter (b) ranges between -0.557 (item 1) and 0.236 (item 5) which implies that item 1 is the easiest and item 5 is the most difficult. Item 1 is the reading of 'It is a fish' which was said to be the easiest, while item 5 reads 'It is a window'. This might be due to the fact that item 1 was used as a guide, while the researcher led in the reading of the first three words. Observation on the field revealed that 90% of the examinees could not read item 5 that is, "It is a window" The reason may be that many examinees could not identify alphabet 'W' and they may likely find it difficult to identify and read word that contains 'W'. Implication is that teachers should pay more

attention to the stimulation of the pre-school children to the learning of the alphabet 'W' and objects that has this letter 'W' in its spellings or pronunciation. This agrees with the results of the reading test administered on 4th grade children in the United States, only 33% performed at or above the proficient level while 33% performed below the basic level (National Reading Panel, 2009).

Table 4.7.3: Difficulty Level of Picture Reading

ITEM	IRT	
	a	**b**
RP01	1.971	0.407
RP02	0.972	-0.853
RP03	2.973	0.455
RP04	3.277	0.527
RP05	1.311	0.123
RP06	1.164	0.764

Table 4.7.3 shows difficulty levels of all the items contained in child's ability to read picture of the Developed ERLT. These were calibrated using 2-PL. The item difficulty parameter (b) ranges between -0.85 (item 2) and 0.76 (item 6) which implies that item 2 "The boy is feeding/playing with the cat" is the easiest and item 6 "The boy is fetching/pouring water" is the most difficult.

Table 4.7.4: Difficulty Level of Recognition of Signs and Symbols

ITEM	IRT	
	a	**b**
RSS01	2.949	0.466
RSS02	3.782	0.454
RSS03	3.595	0.448

Table 4.7.4 shows difficulty levels of all the items contained in child's recognition of signs and symbols of the developed ERLT were calibrated using 2-PL. The item difficulty parameter (b) ranges between 0.45 (item 3) and 0.47 (item 1) which implies that item 3 "Green means Go" is the easiest and item 1 "Red means STOP" is the most difficult.

Table 4.7.5: Difficulty Level of Colour Identification

ITEM	IRT	
	a	**b**
CI/01	0.881	-0.481
CI/02	1.051	-0.327

It can be observed that only two items out of four items contained in the identification of colour subset of DEV-ERLT survived and their difficulty level ranges between -0.33 to -0.48 respectively. That is, identification of blue colour and red colour. It was generally observed on the field that examinees found it difficult to identify the primary colours, this means the items are too difficult for the examinee's age. It was also observed that many of the centres were not colour print rich not to talk of the caregivers exposing the children to these colours in the centres and their immediate environment. The implication of this is that caregivers should make the centres colorful with different prints and objects in order to allow to children learn colours from the embodied environment.

4.8 Research Question Eight: Are there normative data developed to facilitate the interpretation of the Early Reading Literacy Test (ERLT) scores with respect to age, gender, school location and type of school?

To answer this question, stannine and t-score were used to establish the interpretation of ERLT scores. First, examinees abilities were converted to t-scores, then ranked. After which, stannine method was conducted to scale examinees' test scores. Stannine are integers and can be used to convert a test score into a single digit. Perusal of literature has shown that there are very few tests that use stannine to report testing information. However, its scores are typically used with standardised testing and are often reported on the results along with raw scores, these scores can be useful in understanding a relative range of a performance of the examinees. There are 776 examinees scores obtained from a population that is, normally with $\bar{X}$ of 51.06 and a standard deviation of 10.99. The first step is to rank the examinees scores in ascending order and allocate the Stannine scores. Table 4.8 presents examinees' scores and their frequencies

Table 4.8: Examinees Scores and their Frequencies

Examinees Scores	Frequencies
20-25	1
26-31	34
32-37	149
38-43	8
44-49	126
50-55	207
56-61	8
62-67	242
Above 67	1

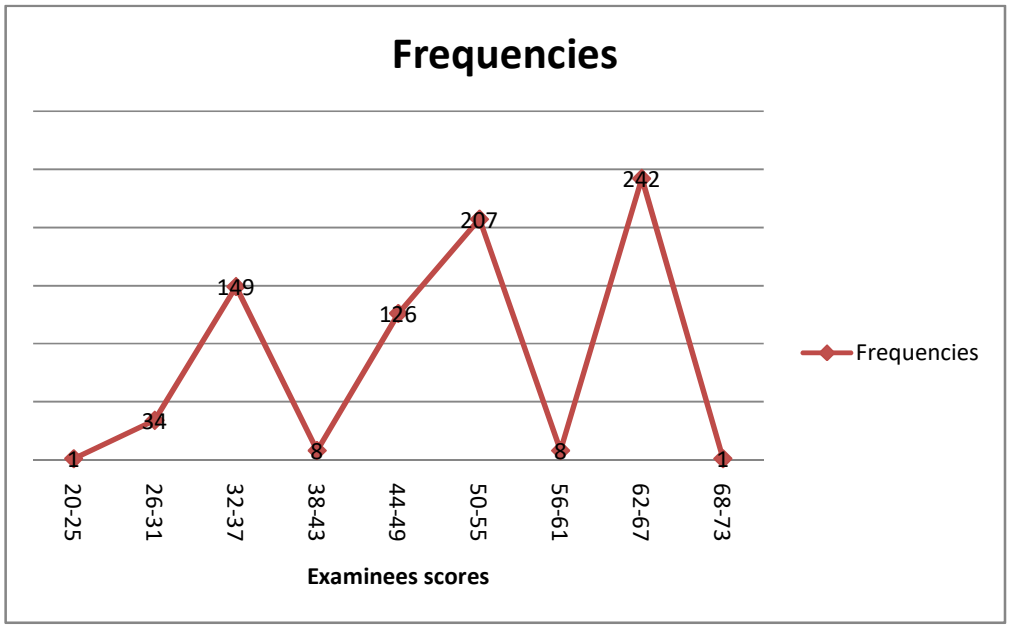

Figure 4.5: Normal distribution of ERLT scores

Figure 4.6 is the graphical illustration of examinees scores and the stannine. That is, examinee scores within 20-25 was given stannine score of 1, while scores within 56-61 carried a stannine score of 7 and so on.

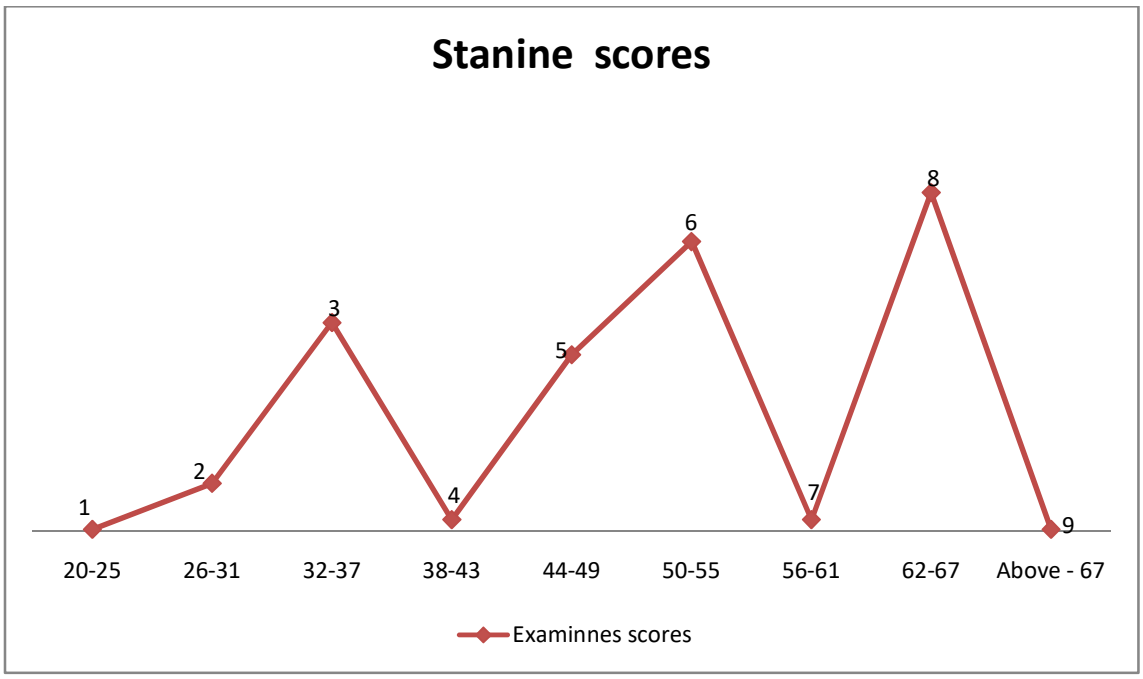

Figure 4.6: Stannine scores of ERLT

Procedure to determine which raw scores become which stannine score or number on the scale is as follows: The first 4% of ranked scores (20-25) was given a stannine score of 1, the next 7% of ranked scores (26-31) was given a stannine score of 2, next 12% of ranked scores (32-37) was given a stannine score of 3, next 17% of ranked scores (38-43) was given a stannine score of 4, middle 20% of ranked scores (44 – 49) was given a stannine score of 5, next 17% of ranked scores (50-55) was given a stannine score of 6, next 12% of ranked scores (56 -61) was given a stannine score of 7, next 7% of ranked scores (62-67) was given a stannine score of 8, while the last 4% of ranked scores (above 67) was given a stannine score of 9.

Consequently, stannines of 1, 2, and 3 reflected below-average achievement in the ERLT compared to the norm group, stannine of 4, 5, and 6 reflected average achievement in the ERLT compared to the norm group, and stannine 7, 8, and 9 reflected above-average achievement in the ERLT compared to the norm group. Succinctly, an examinee who achieved a stannine score that was below average in a particular test, revealed an area in which the test taker needs improvement. If the test taker achieved an average stannine score, the test indicated that he or she performed at about the same level as other examinees who took the test. If the test taker achieved a stannine score that is above average, the test results mean that he or she performed better in that area than other examinees who took the test. Therefore, stannine enables teachers, parents, learners, caregivers and school administrators to have a clearer view of examinees' performance based on the class average performance. Therefore, examinees scores have been transformed to a nine-point scale as shown figure 4.6, which made interpretation of ERLT performance score meaningful.

What are the Normative data established for the purpose of score interpretation along:

i. Age

ii. Gender

iii. School Type

iv. School Location?

Normative Data

Table 4.9: Age Norm

Age	Mean	N	Std. Deviation
Below 4 years	57.85	22.00	10.30
4 years	50.97	197.00	11.49
5 Years	50.83	557.00	10.78
Entire participants	51.06	776.00	11.00

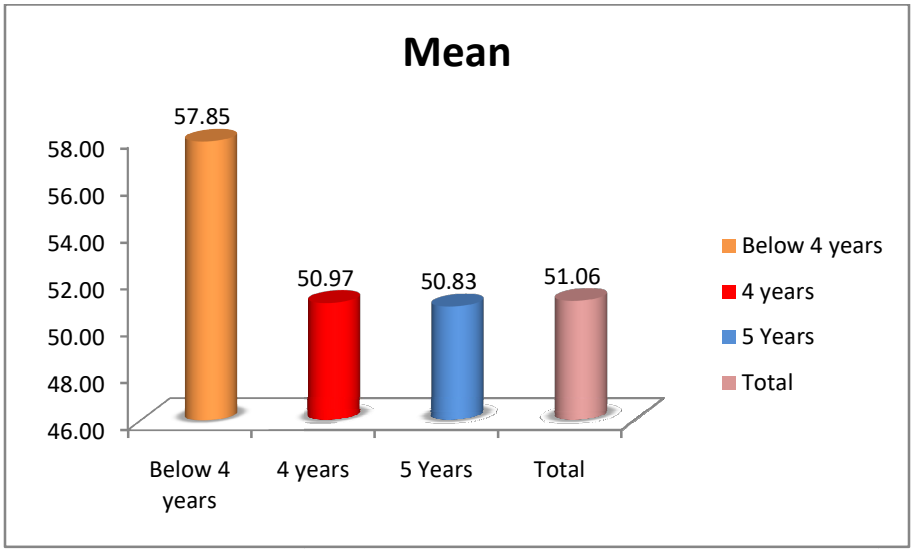

Figure 4.7: Age Norm

Discussions

The data in Table 4.9 and Figure 4.7 show the normative data of examinees by age provides the mean scores of different age groups. It is essential to note that pre-school children with highest age got the lowest mean. The reason for this might be due to the age at which many of these pre-school children enroll into the public schools. From the observations on the field, majority in this category were found in

public pre-schools where there are some factors that may negatively affect their performances.

Mostly in the public schools, children are allowed to enroll into pre-primary classes between age 4+ and 5 years unlike in private sector where there is provision for them from birth. Factors like un-child friendly environment, lack of play materials and colourful print - environment and materials that can stimulate these children to learning especially at the public centres could have contributed to their low performance in the ERLT.

Also, late enrolment in the pre-school centres could contribute to these low performances because the earlier the children are stimulated to learning with appropriate materials and conducive environment, the higher the learning achievement. This supports Sara (2017) who submitted that children who are enrolled in ECCDE centres early between two and a half year of age show high cognitive ability.

Table 4.9.1: Gender Norm

Gender	Mean	N	Std. Deviation
Male	51.79	368.00	10.94
Female	50.40	408.00	11.02
Entire participants	51.06	776.00	11.00

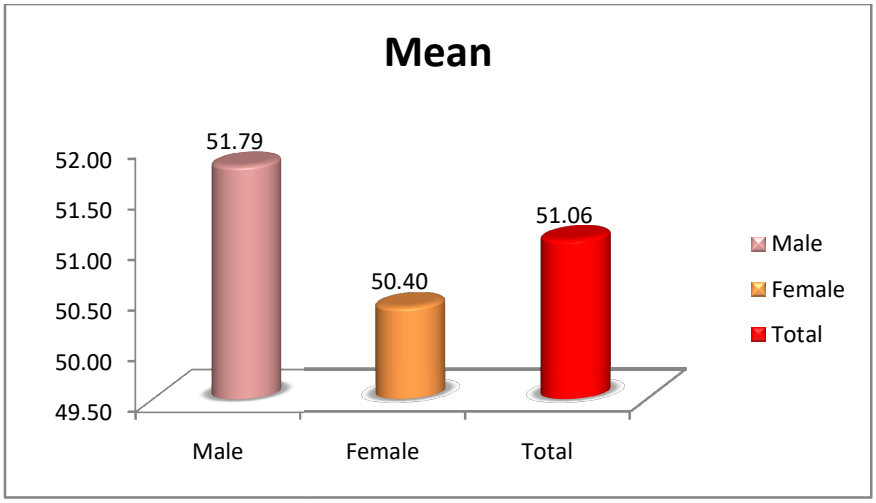

Figure 4.8: Gender Norm

Discussions

The data in Table 4.9.1 and Figure 4.8 show the gender norm. The mean forms the bases for interpreting male and female examinees' scores obtained on the scale. From Table 4.9.2 and Figure 4.8, it could be deduced that male examinees got the mean higher than the total and female examinees' mean. This result negates submission of some researchers like Ijaiya (2007) and Lynn (1994) that females examinees' performed significantly better than males on word fluency. It also contradicts Sara (2017) who found that in spelling performance of examinees in grade one through six, girls scored significantly higher in the six geographical areas of United States. This could be because 99% of the pre - primary teachers are females and at this age, male children are generally more likely closer to female adults. Observation on the field revealed that more males were responding correctly to the items than females, this corroborates with the result. The result also disagrees with the result of the assessment of 23,040 pupils from 960 schools across the nation on the level of competency of primary 4 pupils in numeracy, literacy and life skills which revealed that there was a very little difference in the mean score of girls- 25.8% and the boys - 24.8%. (Falayajo, Makoju, Okebukola, Onugha and Olubodun 1997)

Table 4.9.2: School Type Norm

School Type	Mean	N	Std. Deviation
Private	53.58	596.00	10.45
Public	42.73	180.00	8.36
Entire participants	51.06	776.00	11.00

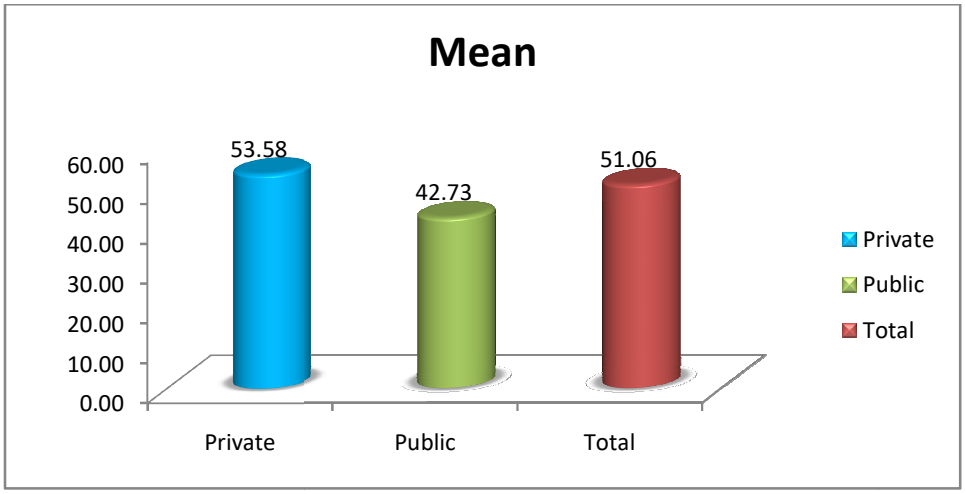

Figure 4.9: Private and public examinees

Discussions

Table 4.9.2 and Figure 4.9 show that the mean scores of examinees from private schools are greater than that of the public and the total. This probably might be because of availability of educational toys, picture books, child-friendly environment and adequate monitoring of the caregivers in many of the private ECCDE centres sampled. The experience in many of the public centres were not the same, many centres lacked educational toys, picture books, while many centres were overcrowded and the environment is not all that child-friendly. This finding corroborates Hartas (2011) who found that literacy rich environment where pre-school children have access to books and other print materials contribute positively to the child literacy and language development. The researcher also observed in the study that generally, more private schools have literacy rich environment than public schools examined. The findings also agree with Sara (2017) who found that play materials and print rich environment can influence children's language skill development.

Table 4.9.3: School Location Norm

School Location	Mean	N	Std. Deviation
Urban	54.28	579.00	9.12
Rural	41.61	197.00	10.62
Total	51.06	776.00	11.00

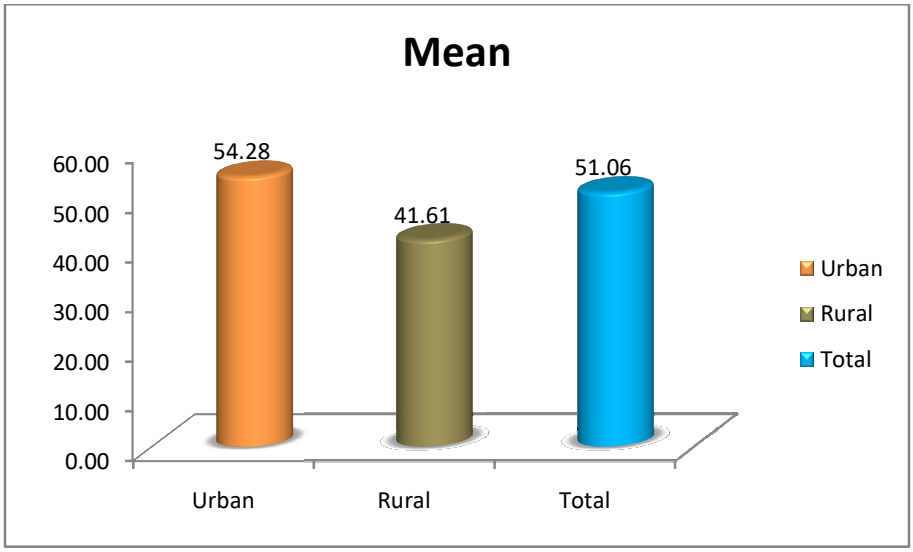

Figure 4.10: School Location Norm

Discussions

The result in Table 4.9.3 and Figure 4.10 indicates that examinees from urban centres have the highest mean scores compared with those from the rural and that of the total mean. The researcher observed that many of the centres in the rural areas were really lagging behind in the availability of learning and play facilities, literacy-rich and conducive to learning environment. Also, it was observed that many centres in the rural areas did not have caregivers that are trained and handle only the pre-schoolers at the public centres, while many of the private centres in the rural areas were substandard with in-experienced and young individuals that just completed their secondary school education. The quality of teachers' factor cannot be overlooked when discussing academic achievement. It was also observed that many schools in the rural areas had only the head teachers and sometimes one additional teacher who handle other classes at the primary section. From experience, many qualified caregivers whose capacities have been developed resist posting to the rural areas and few that were deployed to the rural areas as they were employed, do not stay in the rural areas but go back to the city where they reside after each day's work and sometimes absent from schools. These and some other factors do affect the academic achievement of learners in the rural areas.

The result supports Johnson (2011) who declared that the differences in academic achievement due to location could be as a result of preference by teachers to work in some locations than in others. The researcher concludes that highly qualified teachers prefer to serve in urban areas than in the rural areas. Many teachers do not accept posting to rural areas and even if they accept the posting, they do not live in those rural areas, thereby not being totally committed to their duties. The result of this study also corroborates Williams (2005) who submitted that many rural schools do not have adequate amenities and facilities which may negatively affect the learner's achievement.

CHAPTER FIVE
SUMMARY, CONCLUSION AND RECOMMENDATIONS

5.1 Summary of findings

This study developed and standardised Early Reading Literacy Test (ERLT) for pre-school children. The major findings are;

(1) The data obtained fits 2-parameter logistic model being the lowest value among the models the data were subjected to.

(2) The discrimination index of DRA-ELRT ranged between 0.12 and 6.36, while the difficulty parameter was between -7.92 and 0.83.

(3) A total of 163 out of 226 DRA-ERLT items were assembled as Developed Early Reading Literacy Test (DEV-ERLT) having satisfied the criteria set for IRT framework.

(4) A total of 16 DRA-ERLT items show DIF in gender, 70 items show DIF in school type, while 62 show DIF in school location. However, these items were not deleted because qualitative analysis revealed no incidence of bias in the items.

(5) The estimate of all item parameters in the DEV-ERLT show that the item discrimination value ranges between 0.13 and 6.35 with MEAN=2.57 and SD=1.34 and item difficulty parameter ranges between -6.90 and 0.95 with MEAN=0.81 and SD=1.06. It can be deduced from the analysis that DEV-ERLT discriminates better among the examinees of low and high ability.

(6) The descriptive statistics of ability of test scores ranged from 19.80 to 71.70 with a MEAN of 51.10, which indicates many of the examinees performed a little above average.

(7) The reliability co-efficient of DEV- ERLT sub-sets ranged from 0.62 to 0.94 with sub-set identification of letters having the highest reliability

(8) The difficulty level of sub-sets (a) identification of objects was between -6.90 and 3.99. (b) reading fluency ranged between -0.56 and 0.24 (c) picture reading was between 0.85 and 0.76 (d) signs and symbols was between 0.45 and 0.47, while (e) colour identification ranged between -0.33 and -0.48

(9) Normative scores were based using Stannine and t-score. Stainnes of 1, 2, and 3 that comprised 23% of examinees reflected below average performance in the ERLT compared to the norm group. Stainnes of 4, 5 and 6 that comprised

54% of examinees reflected average while, 7, 8 and 9 reflected above average that is 23% of the examinees compared to the norm group.

(10) Pre - school children of highest age got the lowest mean. Also, the normative data for gender revealed that male examinees got the mean higher than the total and female examinees means.

(11) The mean score of examinees from the private centres is more than the total and that of examinees from the public centres.

(12) The normative data on school location revealed that mean of examinees from urban centres is greater than total and that of examinees from the rural centres.

5.2 Implications of the Study

The stimulation of pre-school children to learning and identifying of English language alphabets in an embedded environment and practical ways becomes important because from the study it was observed that many pre-school children were having difficulty in identifying and differentiating some letters like W, M, B, D, V, X and S. The curriculum developers and caregivers should therefore, device methods of improving this early reading literacy sub-skills in pre-school children. The use of items whose robust psychometric properties are based should be encouraged and promoted at the pre-school level. Standardised testing when combined with other forms of assessment at this level will go a long way in assisting relevant stakeholders in making valid and reliable decision on pre-school children.

5.3 Conclusion

This study developed and standardised test instrument on Early Reading Literacy for pre-school children. Based on the findings, there is a need for stakeholders to make available picture books, story books for pre-schoolers in order to improve their skills in picture reading and reading fluency. Caregivers should endeavour to make the learning centres colour rich so as to help the preschoolers in identifying different colours in their environment.

Stakeholders need to embrace and encourage the use of the standardised instrument in assessing the Early Reading Literacy skills of pre-schoolers so as to have a uniform and valid means of assessing them and arrive at a reliable, appropriate decision and provide the needed intervention at this sub-sector of education.

5.4 Recommendations

The effectiveness and efficiency of a standardised instrument depend on its usage. Therefore, the following recommendations are made based on the findings of the study;

1. Executive summary report of this study should be forwarded to the Universal Basic Education Commission, State Universal Basic Education Board, Ministry of Education who are all in charge of public and private pre- school education and recommend therein that;

(a) the developed and standardised instrument in the study should be adopted for use in replacement of non-uniformed, age and curriculum inappropriate items presently in use at this level of education in our schools.

(b) caregivers at the pre-school level should be trained on the use of the Early Reading Literacy test developed in this study.

(c) caregivers should be encouraged to select items from the instrument in assessing their pre-school children at the middle of third term so as to find out where the children are lagging behind and improve the stimulation on such sub-skills for better performance.

(d) the scores of the pre-school children from the standardised instrument could also be used along with the results from the assessment of other domains in placement of pre-school children at the primary level.

(e) the study revealed that pre-school children in the private schools performed better than those in public, this bring about the recommendation that UBEC/SUBEB should endeavour to improve on the provision of adequate and appropriate educational toys and materials like big books, picture books, literacy rich environment that can stimulate children to acquiring and developing of early reading literacy skills.

(f) UBEC and Ministry of Education should ensure that pre-school children are assessed with items that are curriculum and age - appropriate and also endeavor to see that the psychometric properties of these items are established. This could be better achieved when the establishments collaborate with the experts in training the caregivers.

5.4 Limitations of the Study

Some of the proprietors of private schools that were among the proposed sample of the study were reluctant in accepting that their pre- school children to participate in the study. They were not convinced that it is just for the research purpose. The Local Inspectors of Education in charge of the concerned schools were contacted to assist in educating them better. Though, some of these schools eventually participated, the researcher substituted some with another schools of similar characteristics.

5.6 Suggestions for further Studies

(1) The study developed and standardised instrument for assessment of Early Reading Literacy skills of pre-school children in Nigeria, similar research can be carried out on the numeracy skills of pre-school children in Nigeria.

(2) It was observed in the course of this study that about 59% of pre-school children tested had difficulty in either identifying, differentiating or pronouncing some alphabets like V, W, M, S, X, B, D and the rest. Research on screening assessment on the identification of alphabets and their phonics can also be carried out as a follow-up of this study.

(3) Further studies could be carried out to know the real cause of DIF noticed in the study especially in the performances of rural/urban and public/private schools.

(4) Comparison of the pre-school children early reading literacy performances across the states of the federation could also be conducted using the developed and standardised instrument in this study.

REFERENCES

Aaron, C. C., Nathaniel, O. D. and David, S. T. 2005. *Similarity and discrimination in classical conditioning.* Long Grove, IL: Wave Land Press.

Abarghole, H.M., Khamiripoor, Hosseini, H., Esmaeili, B., Abarghole, J.M. and Abarghole, H.M. 2012. Development and standardisation of achieving test. *Journal of American Science* 8.4:166-171.

Abedi, J., Leon, S. and Mirocha, K. 2001. *Impact of students' language background on content-based performance: analysis of extent information.* Los Angeles: University of California, National Centre for Research on Evacuation, Standards, and Learner Testing.

Abedi, J. and Lord, C. 2001. The language factor in mathematics tests. *Applied Measurement in Education.* 14:219-234.

Abedi, J. 2002. Standardised performance tests and English language students: psychometrics issues. *Educational Assessment* 8.3:231-257.

Adegoke, B. A. 2013, comparison of item statistics of physics performance test using classical test and item response theory framework. *Journal of Education and Practice* 4.22:87-96.

Aiken, L.R. 1979. *Psychological testing and assessment.* Boston: Allyn and Bacon, Inc.

Airassian, M.O. 2001. *Classroom assessment.* 4th Edn. New York: McGraw Hill.

Akinbote, O., Oduolowu, E. and Lawal, B. 2001. *Pre-primary and primary education in Nigeria: a basic text.* Ibadan: Stirling-Horden Publishers.

Akindele, B.P. 2004. The development of an item bank for selection test into Nigerian universities. An expository study. Ph.D. Thesis, Department of Guidance and Counselling, University of Ibadan, xv+210pp.

Akinpelu, J.A. 1987. *Inaugural relevance in education*. A lecture delivered at University of Ibadan. Ibadan: University Press.

Algumalai, S. and Curtis, D.D. 2005. Classical test theory in Slagumali, D. D., Curtis and N. Hungi (eds) *Applied Research Measurement, A book of exemplars*. The Netherland Springer. 1-14

Allen, M.J. and Yen, W.N. 2002. *Introduction to measurement theory.* Long Grove, IL: Wave Land Press.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (AERA, APA, NCME) 1999. *Standards for Educational and Psychological Testing.* Washington, DC: AERA.

Angoff, W.H 1993. S*cales, norms and equivalent scores.* Educational Measurement Washington D.C, American Council on Education

American Guidance Service 1997. *Peabody individual performance test* – revised. Circle Pines, MN: Author.

American Psychological Association 1996. *Standards for educational and psychological tests and manuals.* Washington DC.

Baghaei, P. 2008. The Rasch Model as a construct validation tool. *Rasch Measurement Transaction* 22.1:1144-1146. Available http://www.rasch.org/rmt/rmt221a.htm

Bagnato, S. L., Neisworth, J. T. 2010. Linking authentic assessment and early childhood interventions: Best measure for best practice (2nd ed.) Baltimoo MD: Brooks Publishing Company.

Baker, F.B. 2001. *The basic of item response theory.* 2nd ed. Washington DC: ERIC Clearing House.

Bare, A 2006. Simplified calibration interval analysis. A paper delivered at the Workshop and Symposium under contract with the office of scientific and technical information.

Barnes, B.J. and Wells, C.S. 2009. Differential item functional analysis by gender and race of the national doctoral programme survey. *International Journal of Doctoral Studies*, 4: 77-96.

Barrs, M., Ellis, S., Hester, H., and Thomas, A. 1989. *The primary language record: Handbook for students*. London: Inner London Education Authority/Centre for Language in Primary Education.

Bayley, N. 2006. *Bayley scales of infant and toddler development administration manual.* Harcout Assessment Sa Anotonio, Tx.

Beaton, A. and Aken, N. 1992. Interpreting scales through scale anchoring. *Journal of Educational Statistics* 17.2:191-204.

Bedford, H., Walton, S., and Ahn, J. 2013. *Measures of child development: a review centre for paediatric epidemiology and biostatistics*. UCL Institute of Child Health.

Bishop, S.L. 2011. Convergent validity of the Muller Scales of early learning and the differential ability scales in children with autism spectrum disorders. *American Journal of Intellectual and Development Disabilities* 116.5:331-343.

Block, J.H. 1971. *Introduction mastery learning: Theory and practice* Ed. J.H. Block New York: Holt, Rinehart and Winston.

Bloom, B.S., Hastings, J.T. and Madaus, G.F. 1971. *Handbook on formative and summative assessment of students learning.* New York: McGraw Hill Press.

Boehm, A.E. 2000. *Boehm test of basic concepts.* 3rd Edn. San Antonio. TX: Psychological Corps.

Bowles, S. and Gintis, H. 2002. The inheritance of inequality. *Journal of Economic Perspectives* 16.3:3-30.

Bowman, B.T. and Burns, M. 2001. Eager to learn. *Educating our preschoolers*. Washington D.C.: National Academy press.

Breadly-Johnson, S. 2001. Cognitive assessment for the youngest children: A critical review of tests. *Journal of Pseudo Educational Assessment,* 19.1:18-44.

Braken, B. 2007. Development and use of diagnostic tests to evaluate students. *Journal of Psychology* 5.3:351-364.

Bredekamp, S. 1997. NAEYC issues revised position statement on development of cognitively appropriate practices in early childhood programmes. *Young Children* 52.2:34-40.

Broberg, A.G. 2007. Effect of child care on development of cognitive abilities *Develop Psychology*. 336, 62- 69

Brooks-Gunn, J. and Greg, D. 1997. The effects of poverty on children. *Future of Children* 7.2:55-71.

Broso, W.G. and Brozo, C.C. 1994. Literacy assessment in standardised and zer-failure contexts. *Reading and Writing Quarterly: Overcoming Learning Difficulties* 10:189-200.

Brown, E.G. 1983. *Principles of educational and psychological testing* (3rd Ed.). New York: CBS College Publishers.

Burrill, L.E. 1980. *How a standardised performance test is built. Test service notebook 125.* New York: Psychological Corp.

Cabell, J.., Logan, P. and Knold, T. 2013. Empowering learner independent reading. *Journal of the Association for Childhood Education International* 90.1 Washington, D.C.

Camilli, G. 2006. *Complementary methods in education research.* Washington D.C.: American Educational Research Association.

Camilli, G. and Shepard, L. 1994. *Methods of identifying biased items.* Thousand Oaks, CA Sage publication.

Centre, U.O. C.T.P 2012. *Bayley Scales of Infant Development*. 3rd Edn. Available from: http://www.psychometrics.cam.ac.uk/page/92/bayley-scales.htm

Chall, J.S. 1993. Children's Literacy development. A longitudinal analysis. *Journal of Educational Psychology*, 100: 252 – 271.

Chandra, R. K. 2009. Growth and development. *Indian Journal of Pediatrics* 71: 387-396.

Christy L. Lieras 2008. Do skills and behaviours in high school matter? The contribution of non-cognitive factors in explaining differences in educational attainment and earnings.

Clay, M. 1987. Learning to be learning disabled. *New Zealand Journal of Educational Studies* 22.1:35-38.

_____1991. *Becoming literate: the construction of inner control.* Auckland, NZ: Heinemann.

_____ 1993. *An observation survey of early literacy performance*. Auckland, NZ: Heinemann.

_____1998. *By different paths to common outcomes*. New York, ME: Stenhouse.

Courville, T. G. 2005. An empirical comparison of item response theory and classical test theory item/person statistics. Unpublished doctoral dissertation, Texas A and M University. Retrieved February 9, 2017 from htt://txspace.tamo;ed/bitstream/handle/1969.1/1064/etd-taum-2004B-EPSY-courville-2.pdf

Cronbach, L. J. 2000. Coefficient alpha and method, the internal structure of tests. *Psychometric* 16:297-334.

Cohen, B.I. and Swerdlik, M.E. 2002. *Psychological testing and assessment*. New York: McGraw-Hill Companies. Inc.

Coles, G. 1987. *The learning mystique*. New York: Ballantine.

Cosgrove, J., Shiel, G., Sofroniou, N., Zastrutzki, S. and Short, F. 2005. Education for life: the performance of 15-year olds in Ireland in the second cycle of PISA. Dublin: Educational Research Centre.

Cress, S.W. 2004. Assessing standards in the 'real' kindergarten classroom. *Early Childhood Education Journal* 32:95-99.

Crocker, L. and Algina, J. 1986. *Introduction to classical and modern test theory*. Belmont, CA: Wadsworth Press.

Crooks, T.J., Kane, M.T. and Cohen, A.S. 1996. Threats to valid use of assessment. *Assessment in Education* 3:265-285.

Crouch, S. and Skoog, D. A. 2007. *Principles of instrucognitive analysis.* Pacific Gove: Brooks Cole. ISB4.

Davidson, K. 2010. The integration of cognitive and socio-cultural theories of literacy development: why? How? *The Alberta Journal of Educational Research 56.3:246-256.*

Demars, C. 2010. *Item response theory. Understanding statistics measurement.* Oxford: Oxford University Press.

Doig, K.B. 1999. The child development inventory: A develop cognitive outcome measure for follow-up of the high-risk infant. *The Journal of Paediatrics* 135.3:358-362.

Dodeen, H. 2004. The relationship between item parameters and item fit. *Journal of Educational Measurement* 41:261-270.

Dorans, N. J. and Holland, P. W. 1993. Detection and description: Mantel-Haenzel and standardisation Eds. Differential item functioning (pp. 35-66). Hillsdale, NJ, US: Lawrence Erlbaum Associates Inc.

Dowing, S. and Haladyn, T.M. 2006. Handbook of test development. Mahwah, N.I. Lawrence Erlbaum Associates.

Eigsti, I. M. 2013. Peabody picture vocabulary test. In Volkma, F. R. Eds. *Encyclopedia of Autism Spectrum Disorder*. New York, NY: Springe.

Duncan, G. S. and Jeanne, B. G. 1997. Income effects across the life-span: integration and interpretation. In *Consequences of growing up poor.* G.J. Duncan and Jeanne Brooks-Gunn Eds. New York: Russell Sage Foundation.

Duong, M. 2004. Introduction to item response theory and its application. New York: Harcourt Brace. http:www.msu.edu/-dwong/learnerworkarchive/cep900f04-rpd/mihn-itemresponsetheory…16/5/2015.

Dura-Bellat, M. and Suchaut, B. 2005. Organisation and context, efficiency and equity of educational systems. What PISA Tells us. *European Educational Research Journal* 4.3:181-94.

Duyne, M. 2011. French norms and validation of the child development inventory (CDI). *Clinical Paediatrics* 50.7:636-647.

Ehri, L.C. 2006. Development of sight word reading: stages and findings. In M.J. Snowling and C, Hulme Eds. *The science of reading: A handbook* (pp. 135-154). Malden, M.A.: Blackwell.

Eigsti, I. M. 2013. Peabody picture vocabulary test. In Volkma, F. R. Eds. *Encyclopedia of Autism Spectrum Disorder*. New York, NY: Springe.

Eluwa, A.N., Idowu, E.O. and Abang, B.K. 2011. Assessement of Mathematics performance test: a comparison between classical test theory and item response theory (IRT). *Journal of Education and Social Research* 1.4:99-106.

Enu, V. O. 2015. Validation and calibration of mathematics and geography items for joint command schools promotion examination in Nigeria using Item Response Theory. Ph.D. thesis, International Centre for Educational Evaluation, University of Ibadan, Nigeria.

Enuwah, C.I. and Akwa, A.M. 2014. Development and standardisation of performance test in senior secondary school mathematics using item analysis theory. Paper presented at the 6th International Conference on Education and New Learning Technologies, Spain.

Essen, C.B. 2015. Differential item functioning of 2014 unifield tertiary matriculation test mathematics of candidate in Akwa Ibom state Nigeria. Unpublished Doctoral Dissertation, University of Calabar Nigeria.

Fakayode, O. T. 2016. Relative effectiveness of classical test and item response theories in equating West African Examination Council mathematics test scores for June and November. Ph.D. research post-field seminar paper presented at the Institute of Education, University of Ibadan.

Falayajo, W., Makoju, G.A.E., Okebukola, P., Onugha, D.C. and Olubodun, J.O. 1997. Assessment of learning performance of primary four pupil in Nigeria: National report. Abuja: Federal Ministry of Education/UNICEF/UNESCO.

Falaye, B. A. and Adefisoye, B. T. 2016. Continuous assessment practice of secondary school teacher in Osun State. *Journal of Psychology and Behavioural Science* 4.17:44-55.

Fan, X. 1998. Item response theory and classical test theory: an empirical comparison of their item/person statistics. *Journal of Educational and Psychological Measurement,* 58.3:1-17.

Feldt, L.S. and Brennan, R.L. 1989. *Reliability educational measurement.* R. Linn Ed. New York: American Council on Education/Macmillan.

Federal Republic of Nigeria 2005. National Policy on Education. 3rd Edition.

Federal Republic of Nigeria 2014. National Policy on Education. 6th edition p. 5.

Frank, H. and Altheon, S.C. 1994. *Statistics: concepts and application*. Cambridge University.

Georgieu, N. 2008 .Item analysis of C, D and E series from Raven's *Standard progressive Matrice with item responses theory two-parameter logistic model*, *Europe, Journal of psychology.*

Gillon, G.T. 2017. *Phonological awareness*. New York, London: Guilford Press.

Glascoe, F. P. and Dworkin, P. H. 1995. Obstacles to effective developmental surveillance. Vanderbilt University School of Medicine, Child Development Center, Nashville, TN 37232.

Glaser, R. 1988. Instructional technology and measurement of learning outcomes. *American Psychologist* 18:519-521.

Glazier, S.M. and Searfos, A.S. 1988. Re-examining reading diagnosis. In S. Glazer, L. Searfoss, and L. Gentile Eds. *Reexamining reading diagnosis: new trends and procedures* pp. 1-11. Newark, DE: International Reading Association.

Gollenberg, A.L. 2010. Concurrent validity of the parent-completed ages and stages questionnaire. 2nd edn. With Bayley Scales of Infant Development II in a low-risk sample. *Child: Care, Health and Development* 36.4:671-676.

Good, R.H. and Kasminski, R.A. 2002. *Dynamic indicators of basic early literacy skills – (DIBELS). Administration and scoring guide*. 6th Edn. Eugeue: Institute for the Development of Educational Performance.

Goodlad, J.I. 1979. *What schools are for?* Bloomington, IN: Phi Delta Kappa Educational Foundation.

Goodman, B.S. 2009. Kidwatching: observing children in the classroom. In A. Jagger and M.T. Smith-Burke Eds. *Observing the language learner* pp. 9-18. Newark, DE: International Reading Association.

Goodwin, W.L. and Goodwin, L.D. 1993. *Measuring young children. Handbook of research in early childhood education.* B. Spodek, Ed. New York: Free Media.

Gosswami, U. 2003. Early phonological development and the acquisition of literacy. In S.B. Neuman and D.K. Dickson Eds. *Handbook of early literacy research* (pp. 111-125). New York, London: Guilford Press.

Granewell, C. 2000. *Sex differences in second language learning.* The University of Victoria.

Green, K.E. and Fraiton, C.G. 2002. Survey development and validation with the Rasch Model. A paper presented at the International Conference on Questionnaire Development. 14-17th November, 2002, Eassessement and Testing. Charlestan. S.

Green, S. B. 2003. A coefficient alpha for test-retest data. *Psychological Method,* 8:88-101.

Greggory, B.O. 1997. Introduction. In E Gregory Ed. *One child, many worlds: early learning in multicultural communities* pp. 1-8. New York: Students College Press.

Groulund, G. 2006. *Make early learning standards come alive: connecting your practice and curriculum to state guidelines.* St. Paul, MN: Redleef Press.

Groulund, G. 2006. *Make early learning standards come alive: Connecting your practice and curriculum to state guidelines.* St Paul, M. N.: Redleaf Press.

Haider, S. I., Asif, S. E. 2011. Quality Control Training manual. Comprehensive Training Guide for API, CRC Press p.49.

Hambleton, R. and Jones, R. 1993. Comparison of classical test theory and item response theory and their application to test development. *Educational Measurement: Issues and Practice* 12: 38-47.

Hambleton, R. K., Swaminathan, H. and Rogers, H. J. 1991. *Fundamental of Item Response Theory.* ISBN Rf. 196N34.

Harold, I. and Francis, P. G. 1995. Assessing children's development using parent's reports. The Child Development Inventor, Manual Minneapolis, Mig. Corp 233 Spring St. New York NY 10013.

Harrison, P. 1990. *AGS Early Screening Profiles (ESP).* Circle Pines, MN: American Guidance Service.

Harvey, R. and Hammer, A.1999. *Item response theory* 27.3;353-383.

Hartas, D. 2011. Families' social backgrounds matters: socio-economic factors, home learning and young children's language, literary and social outcomes. *British Educational Research Journal*, 3.6: 893-914.

Heckman, J., Pinto, R. and Savelyev, R. 2013. Understanding the mechanism through which an influential early childhood program boosted adult outcomes.

Herrnstein, R.J. and Murray, C. 1994. *The bell curve: intelligence and class structure in American life*. New York: Free Press.

Hiebert, E.H. and Callfee, B.AC. 1993. Advancing academic literacy through students' assessments. *Educational Leadership* 50-54.

Hills, T.W. 1992. Reading potentials through appropriate assessment. In S. Breddcamp and T. Rosegrant (Eds.) Reading potential through appropriate curriculum and assessment for young children (pp. 43-64). Washington, DC: National Association for the education of young children.

Hodges, C. 1997. How valid and useful are alternative assessments for decision making in primary score classrooms? *Reading Research and Instruction* 36.2:157-173.

Hogan-Bassey, V. 2011. Detection of differential item functioning in Mathematics multiple choice test administered by National Test Council (NECO) in Abuja Unpublished M.Ed Dissertation, University of Nigeria, Nsukka.

Hohman, M. and Weikart, D. *Educating young children: active learning practices for preschool and child care programme.* Ypsilanti, MI: High/Scope Press.

Holland, P. W. and Thayer, D. T. 1988. Differential item performance and the Mantel-Haenszel procdure. In H. Wainer and H. I. Braum (Eds.), *Test validity* (pp. 129-145). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Hood, M., Conlon, E., and Andrews, G. (2008) Pre-school homes Literacy practices and children's literacy development: A longitudinal analysis. *Journal of Educational Psychology* 100:252-271.

Hopkins, L.T. 1941. *Interaction: the democratic process*. Boston, D.C: Heath.

Huang, F. L., Ford, K. L., Invernizz, M. and Fani, X. 2013. Phological awareness, literacy Spanish. Retrieved from https://eric.ed.govt/

Huff, 2000. Review of items response theory practices in organisation research. Article in organisation research 2017 with 831. Doi 101177/109442811668970

Hutchins, R.M. 1936. *The higher learning in America.* New Haven: Yale U.P.

Igbokwe, E. 2004. Development of Mathematics item bank for placement of students into federal unity schools in Nigeria. Unpublished Ph.D. Thesis, International Centre for Educational Evaluation, University of Ibadan, xiv+157pp.

Ijaiya, B.S. 2007. A psycho-pragmatic description of performance in the English of selected nursery school children in Kwara State. Unpublished Ph.D. Thesis, University of Ibadan, Ibadan, Nigeria.

Ipaye 1996. Future trends in special education. A key note address presented at the CENDP/UNESCO sponsored training programme on current resources in special education at FCE (Special) Oyo.

IRA/NCTE Joint Task Force on Assessment 1994. Standard for the assessment of reading and writing. New York, DE: International Reading Association and Urban, IL: National Council of Students of English.

Ireton, H. 1992. *Child development inventory manual.* Retrieved: 17 September, 2016./www.childdevrev.com/page11/page59/files/cdi-manual.pdf

Ireton, H. and Glascoe, F.P. 1995. Assessing children's development using parents' report. The Child Development Inventory. *Clinical Paediatrics* 34.5:248-255.

Isiugo-Abanihe, I. and Labo-Popoola, O.S. 2004. School type and location as environcognitive factor in learning English as a second language. *West Journal of Education*. 24.1:55-63

Jacob, K., Darrel, S.B. and Paulgros, B. A. 2018. An introduction to item response theory analysis of eating assessment tool. *Journal of the Australian Society for the Study of Brian Impairment.*

Jacobson, S.H.Z. 1997. A comparison of childhood assessment and standardised measure for programme assessment. Unpublished Ph.D. Thesis, Virginia Polytechnic Institute and State University, U.S.A.

Jacoby, R. and Glauberman, P. 1995. *The bell curve debate: history documents opinions.* New York: Random House.

James, J. H., Bat, H. H., Golsteyn, J. E. 2016. What grades and achievement tests measure.

Jenifer, B. 2009. Decomposing the impact of portfolio constraints. MSCI Barrs Research paper No. 209-30.

Johnson, A.B 2011. School mapping and resource supply as correlates of students academic performance in Kwara State Secondary Schools, *Journals of Language Teaching and Research,* 2.5:1052-1060.

Johnston, B. and Johnson, K. A. 2002. Preschool immersion education for indigenous languages: A survey of resources. *Canadian Journal of Native Education*, 26.2:107-223.

John-Steiner, V. and Mahn, H. 1996. Socio-cultural approaches to learning and development: a Vgotskian framework. *Educational Psychologist* 31.3/4:191-206.

Johnson, P., Guice, S., Baker, K., Malone, J., and Michelson, N. 1995. Assessment of teaching and learning in "literature based" classrooms. *Teaching and students education*. 11.4:359-371.

Johnson, S.B. 1998. *Positive assessment of literate activity*. New York: Longman.

_____2011. Assessment as social practice. In D. Leu and C. Kinzer Eds. *Forty-second yearbook of the National Reading Conference*. Chicago, IL: National Reading Conference.

_____1998. *Knowing literacy: constructive literacy assessment.* New York, ME: Stenhouse.

_____1997. The consequence in using standardised tests. In S. Murphy Ed. *Fragile evidence: a critique of reading assessment* pp. 89-101. Mahwah, NJ: Lawrence Erlbaum.

Johnston, P.H. and Rogers, R. 2000. *Assessment of literary development in early childhood: Centre on English learning and performance.* N. Y: Longman.

Juel, C. 1991. The role of decoding in early literacy instruction and assessment. In L.M. Morrow and J. K. Smith Eds. *Assessment for instruction in early literacy* pp. 135-154. Englewood Cliff, NJ: Prentice-Hall.

Kaderavek, J. N. and Pertimonten, K. 2014. Developing reading literacy skills in children. *Journal of the Association for Childhood Education International* 10.6. Washington, D.C.

Kaufman, A. and Kaufman, N. 1983. *Kaufman Assessment Battery for Children (K-ABC).Sampler Manual.* Circle Pines, MN: American Guidance Service.

Kennedy, E. 2003. Raising test scores for all students: an administrator's guide to improving standardised test performance. N.P. Corwin.

Kenneth, N. R. 2004. *Qualitative Research Methods in Educational Planning.* Paris, France: International Institute for Educational Planning/UNESCO.

Kirabo Jackson C. 2012. Non-cognitive ability, test scores and teacher quality: Evidence from 9th grade teachers in North Carolina.

Lareau, A. 1989. *Home advantage: social class and parental intervention in elementary education.* London: Falmer Press.

_____1998. *Meaningful literacy: constructive literacy assessment.* New York, ME: Stenhouse.

_____2003. *Unequal childhoods: class race and family life.* Berkeley: University of California Press.

Levitov, R. 2015. Item analysis of students comprehensive test for research in teaching. *International Journal of Science and Research (IJSR)*, (12) B Volume 4 Issue 5.

Li Yamei, Cohen Allens and Ibara Robert, A 2004. *Characteristics of Mathematics Items Associated with gender, DIF. International journal of testing 4(2.115 – 136).*

Linacre, J.M. 2004. Test validity and Rasch measurement: construct, content, etc. Rasch Measurement Transaction 18.1:920-971. Available htpp:///www.rasch.org

Linn, R.L. and Muller, M.D. 2005. Measurement and assessment in teaching (9th Edn.) Upper Saddle River, NJ: Merill/Prentice-Hall.

Lynn, O. R. 1994. Sex differences on the differential intelligence FEM. test in British and American Adolescents. *Educational Psychology* 12:10 – 100.

Lord. F.M. 1992. Sex differences on the differential intelligence test in British and American adolescents. *Educational Psychology* 12:101-106.

Lord, F. M. 1970. Item characteristic curves estimated without knowledge of their Mathematical form. A confrontation of Birn Baumbo logistic model. *Psychometrika* 35 : 43 – 50.

Madaus, G., Russell, M. and Higgins, J. 2009. *The paradoxes of high stakes testing: How they affect students, their parents, students, principals, schools and society*. Charlotte NC: Information Age Publishing.

Madu, B.C. 2012. Analysis of gender-related differential item functioning in mathematics multiple choice items administered by West African Test Council (WAEC). *Journal of Education and Practice*, 3.8:71-79.

Mann, V., Sassanuma, S., Sakuma, N., Mabaki, S. 1990. Sex differences in cognition abilities: a cross-cultural perspective. *Neuropsychologia* 28:1063-1077.

Mardell-Czundowski, C.D. and Goldenberg, D.S. 1998. *Developcognitive indicators for the assessment of learning.* 3rd Edn. Circle Pines, MN: American Guidance Service.

MccAfee, A., Leong, D.J., and Bodrova, E. 2004. *Basics of assessment a primer for early childhood education*. 4.7:85-95. Washington, D. C.: National Association for the Education of Young Children.

McCarthy, D. 1983. *McCarthy scales of children's abilities.* New York: Psychological Corp.

McGuiness, E. 1998. *Our children can't read.* New York, Toronto, ON: Free Press.

McMaken, J. 2003. Early childhood assessment. Denver, Co: Education Commission of the United States https://www.ecs.org/html/document

McNamee, G.D. 1995. A vygotskian perspective on literacy development. *School Psychology International* 16:185-198.

Meisels, S.J., Marsden, D.B., Wiske, M.S. and Henderson, I.W. 1999. *Early Screening Inventory – Revised (ESI-R)*. New York: Pearson Education.

Mercer, C.D. 1997. *Learner with learning disabilities* (5th Edn) Upper Saddle River, NJ: Prentice-Hall.

Merrell, K.W. 2002. *Preschool and kindergarten behaviour scale.* 2nd Edn. Austin, TX: PRO-ED.

Messick, S. 1989. Meaning and values in test validation. *The Science and Ethics of Assessment. Journals*, Sage Pub. Comodo

Metibemu M.A 2016 Comparison of classical test and item response theories in the development and scoring of senior secondary school physics tests in Ondo state, Nigeria. A PHD research post field seminar paper presented at the institute of education, University of Ibadan.

Meyer, J.P. and Zhu, S. 2013. Fair and equitable measurement of learner learning in massive open online courses (MOOCs): an introduction to item response theory, scale linking and score equating. *Research and Practice in Assessment* 8: 26-39.

Mfonobong, E. U. 2004. Child abuse and its implications for the education sector in Nigeria. *Journal of Educational Psychology* 150:332-340

Michalak, R. 2014. Environmental factors and literacy learning and instruction. A Master thesis Capstone Project Department of Language, Learning and Leadership. State University of New York at Freedonia, New York.

Miller, L.J. 1993. *First step: screening test for evaluating preschoolers (First Step)*. San Antonio, TX: Psychological Corp.

Moore, D. G., Goodwin, J. E. and Oates, J. M. 2012. A modified version of the Bayley scales of infant development – 11 for cognitive matching of infants with and without Down Syndrome. *Journal of Intellectual Disability Research,* 52.6:554-561.

Morgan, H. 2014. Use of standardised test to develop literacy level of students in primary schools. *American Journal of Educational Research* 2.9:811-816.

Moshtaghian, T.T.S. 2009. Construction and standardisation of performance test. Unpublished M.A. Thesis, Departmenr of Educational Psychology, Tehran Unity, Iran.

Mullen, E. M. 1995. Mullen Scales of Early Learning. https://www.pearsonclinical.com> product.

National Educational Research and Development Council (NERDC) 2005. Integrated early childhood care and development education curriculum. Published by UNICEF, Abuja.

_____2014. Early learning development standards for early child care development education. Published by UNICEF, Abuja.

National Reading Panel 2009. Report of the National Reading Panel: Teaching children to read. Washington, DC. U.S. Department of Health and Human Services.

National Association for the Education of Young Children 1991. Guidelines for the appropriate curriculum content and assessment in programmes serving children ages 3 through 8. A position statement of the National Association for the Education of Young Children and the National Association of Early Childhood Specialist in the State Department of Education. *Young Children* 21-38.

_____1998. Learning to read and write: develop cognitively appropriate practices for young children. A joint statement of the International Reading Association (IRA) and the National Association for the Education of Young Children (NAEYC).*Young Children* 54.4:30-46.

Nenty, H.J. 2004. From classical to item response theory (IRT). An introduction to a desirable transition issues in educational measurement and assessment in Nigeria. In honour of Wole Falayajo Eds. by Omaze A. Afemikhe and J. Gbenga Adewale. Educational Research and Study Group Institute of Education, University of Ibadan, Ibadan, Nigeria.

Newborg, Stock, Wrek, Guidubaldi and Svinicki 1988. *The impact of childcare and parent-child interactions on school.* New York: CBS College Publisher.

Niel, J. D., Tim, P. M. and Dancel, R. E. 2010. Principles and practices of test score equations. *Educational Testing Service (ETS)*, Princeton, New Jersey.

Nihira, K., Lelands, H. and Lambert, N. 2014. Adaptive behaviour scale – Residential and community examiner's manual (2nd ed.) ProEd. *Creative Education* 5.7:28.

Nkpone, H.L. 2001. Application of latent trait model in developing and standardization of physics performance test for senior secondary schools. Ph.D. Thesis, Institute of Education, University of Ibadan.

Nunes, S.R., Willows, D.M., Schuster, B.V., Yaghoub-Zadeh, Z. and Shanahan, T. 2001. Phonemic awareness instruction helps children to read: evidence from the national reading panel's meta-analysis. *Reading Research Quarterly* 36:250-287.

Nunnally. J. C. and Bernstein. I.H. 1994. *Psychometric theory science and education.* New York, N.Y.: M.C. Graw-Hill, Inc.

Nworgu, D.G. 2003. *Educational measurement and assessment theory and practice.* Nsukka: I University Trust Publishers.

_____2011. Differential Item Function: A critical issue in regional quality assurance. A paper presented in NEARN Conference.

Nworgu, D.G. and Odili, J.N. 2005. Analysis of DIF in WAEC SSCE Biology multiple choice test. Review of Education 16.2:140-152.

NYESED 2019. New York State pre-kindergarten foundation for the common core. Retrieved from http//wwwp12ngsedgov/cia/commoncore standard.

Nystraandy, M. 2006. The social and historical context for writing research. In C.A. MacArthur, S. Graham, and J. Fitzgerald Eds. *Handbook of writing research* (pp. 11-27). New York, London: Guilford Press.

Obanya, P.A.I. and Ezewu, E.E. 1981. Socio-psychological background of primary school pupils in two Nigerian cities. Research project on correlates of academic performance in primary schools. International Development Research Centre, Ottawa, Canada.

Odile, J.N. 2009. Effect of language manipulation on differential item functioning in a multicultural setting. A paper presented at the 27th Annual Conference of Association for Educational Assessment in Africa, Cameroon, 2009, August 24-28, 2009.

Odinko, N. M. 2002. Influence of home and School factors on identification and matching skills among pre-primary school Children. Ph.D. Thesis, University of Ibadan, Ibadan.

Oduolowu, E. 2011. *Contemporary issues in early childhood education.* Ibadan: Franco-Ola Publishers.

Ofodu, G. O. and Lawal, R. A. 2010. Effect of Gender and Social Location on Ekiti State Secondary School students. Performance of Reading comprehension in English Language. *Journal of Education and Practice.* 18.5: 2017

Ogbebor, U.C. 2012. Differential item functioning of Economics question paper of National Test Council (NECO) in Delta State. M.Ed Dissertation, International Centre for Educational Evaluation, University of Ibadan, Ibadan, Nigeria.

Ogbebor, U.C. 2016. *Construction of Mock Economics Test for Senior Secondary School Learner in Delta State, Nigeria using classical test and item response. Theories . A. Ph.D. research post Held seminar paper presented at the Institute of Education, University of Ibadan.*

Ohunche, R.O. and Akeju, S.A. 1988. *Measurement and assessment in education.* Ibadan: Aricana-Feb Publishers Ltd.

Ojerinde, D. 2013. Classical Test Theory (CTT) vs Item Response Theory (IRT): an assessment of comparability of item analysis results. A paper presented at the Institute of Education, University of Ibadan. 23rd May, 2013. Lecture Series.

Okediran, A. and Danesty, A.H. 2002. Etiological factors and effects of street working behaviour among Nigerian youth. *Journal of Social Problems*, School of Arts and Social Sciences FCE (Special) Oyo.

Okonkwo, S. C. 2000. Relationship between some school and teacher variable and students achievement in Mathematics. *Journal of Science Association of Nigeria* 35:43-49.

Oliver, S.J. and Klugman, E. 2006. Play and standards-driven curricula: can they work together in pre-school. *Exchange,* 170:12-16.

Onasanya, S. A. and Omosewa, E. O. 2000. Effect of improvised and standard instructional materials on secondary school students' academic performance in physics in Ilorin, Nigeria. *Singapore Journal of Scientific Research*. Volume 1, Issue 1 https://scaternet.

Opasina, C. O. 2009. Development and validation of alternative to practical physics test using item response models. Ph.D. thesis, International Centre for Educational Evaluation, University of Ibadan, Nigeria.

Osadebe, P.U. 2001. Construction and validation of test. Paper presented at the University of Port Harcourt.

_____2014. Test standardisation for assessment and comparing of students' measurement. *International Education Studies*, 7.5:94-103.

_____2015. Assessment score of university lecturers. *Journal of Educational Practice* 5.2:8-14.

Oshima, T.C. and Morris, S.B. 2008. Raju's differential functioning of items and tests (DFIT). Educational Measurement: Issue and practice 27, 43-50.

Osterlinde, S.J. 1989. *Construction test items*. Boston: Khuwer Academic Journal.

Owen, F. and Jones, R. 1994. *Statistics*. London: Longman Group UK Limited.

Pacini – Ketchabaw and Pence (2011) *Journal of the Childhood Education.* Nov/Dec. 2014. Vol.90 no.5.

Pariss, S. Turner and Lawton 1991. A development perspective on standardised performance testing. *Educational Researcher* 20.5:12-20.

Pearson, K. 2012. Baylay Scales of infant and toddler development. 3rd Edn. (Bayley – III). Available from: http://www.pearsonclinical.co.uk/psychology/childgeneralabilities/Bayleyscalesofinfantandtoddlerdevelopmentthirdedition%28Bayley-111%29/forthisproduct/Bayley-IIIUKWalidartion.aspx.

Perez, B. (1998). Literacy diversity and programmatic responses. In B. Perez Ed. *Socio-cultural contexts of language and literacy* (pp. 3-20). Englewood Cliffs, NJ: Erlbaum.

Phelps, R.P. 2012. The effect of testing on learner performance 1910-2010. *International Journal of Testing* 10.10:2012.

Phyye, E.D. 1997. Epilogue: classroom assessment – looking forward. In G.D. Phye Ed. *Handbook of classroom assessment: learning adjustment and performance* pp. 531-534. New York: Academic Press.

Purcell-Gates, V., Jacobson, E. and Degener, S. 2004. *Print literacy development: uniting cognitive and social practice theories*. Cambridge, MA: Harvard University Press.

Progar, S. and Socan, G. 2008. An empirical comparison of item response theory and classical test theory. *Psilhilosra Obzorja/Horizons of Psychology,* 12.3:5-24.

Ray, I and Bryant, P. 1990. Social background, phonological awareness and children's reading British, Journal of Development Psychology, 8: 209-225.

Reardon, S.F. 2011. *The widening academic performance gap between the rich and the poor new evidence and possible explanations.* New York: Rusell Sage Foundation.

Reise, S.P. and Walter, N.G. 2010. Measuring psychopathology with non-standard item response theory model: fitting the four-parameter model to the Minnesota multiphasic personality inventory. In Embretson, S., Ed. *New direction in psychological measurement with model-based approaches.* American Psychology Association Washington, DC, 147-173.

Robins, J., Breslow, N. and Greenland, S. 1988. Estimates of the Mantel Haenszel variance consistent in both sparse data and large strata limiting models. *Bometrics* 42:311-323.

Rogers, H.J. 1991. *Fund cognitive of item response theory*. Newbury Park, CA: Sage Publications, Inc.

Rogof, B. 1995. Observing socio-cultural activity on three planes: participatory appropriation, guided participation, apprenticeship. In *Sociocultural Studies in Mind*. J. Wertsch, P. del Rio, and A. Alverez. Cambridge: Cambridge University Press 139-164.

Ronald J. C. and Merk S. (2010). Psychological testing and assessment: an introduction to test and measurement. Sage Publication Inc.

Roskos, K. and Neuman, S.B. 1993. Descriptive observations of adults' facilitation of literacy in young children's play. *Early Childhood Research Quarterly* 8:77-97.

Roskos, K. and Vukelich, C. 2006. Early literacy policy and pedagogy. In D.K. Dickinson and S.B. Neuman Eds. *Handbook of early literacy*. 2:295-308. New York: Guilford Press.

Rupp, A.A. 2009. Item response modeling with Bilog-MG and multiblog for windows. *International Journal of Testing* 3.4:365-384

Ryan, J. D. K., Osborn P. S. and Rivera, R. 2002. An overview of item response theory, presentation at the Arizonal education research organization conferences Arizona state University west,

Rydz, D. 2006. Screening for delay in the cognitive development in the setting of a community paediatric clinic: A prospective assessment of parent-report questionnaires. *Paediatrics* 118.4:1178-1186.

Sara, R. 2017. School transition and school readiness. An outcome of Early Childhood Development. New York: Vintage Book.

Scott Little, Kagans and Clifford 2013. Early childhood development as economic development. New York: Guilford Press.

Scriven, I. 1999. Nursery children as emerging readers and writers. In R. Campbell Ed. *Facilitating preschool literacy* pp. 169-191. Newark, DE: International Reading Association.

Segal, M. and Webber, N.T. 1996. Nonstructured play observations: guidelines, benefits and caveats. New visions for the develop cognitive assessment of infants and young children. S.J. Meisels and E. Fenichel Eds. Washington D.C: Zero to three: National Centre for Infants, Toddlers and Families.

Shanahan, T. and Lonigan, C. J. Eds. 2013 Literacy in pre-school and kindergarten children. The National Early Literacy Panel and beyond. Baltimore: Brookes Publishing.

Shaywitz, S. 2005. *Overcoming dyslexia*. New York: Vintage Books.

Sheldon, S. 2006. Formalise the informal or 'informalise' the formal. The transition from pre-school to primary. *International Institute for Educational Planning Newsletter* March 24.1: 7,

Shepard, L.A. and Smith, M.L. 1988. Synthesis of research on school readiness and kindergarten retention. *Educational Leadership* 44:78-86.

Shoaga, O. 2017. Nurturing reading proficiency of pupils through phonics. Adopting integrated curriculum for early childhood education: The Nigeria teachers' dilemma (Part 2) https//trulgalirefoundation.org

Shoemaker, B. 1989. Integrative education: A curriculum for 21st century. *Oregon School Study Council*, 33.2:5-8.

Singhanie, R and Sonksei, P., *Indian J. Pediatric* 2004. 71: 387 doi: 10. 1007/Bfo225622

Sofronion, N. and Kellaghan, T. 2004. The utility of TIMSS scales in predicting students subsequent state examination performance. *Journal of Educational Measurement,* 41: 311-329.

Sonnen-Schen, C. K., Stapleton, L. M. and Benson, D. A. 2010. Promoting emergent reading literacy. *Journal of the Association of Childhood Education International* 70.4. Washington, DC.

Sparrow, S.S., Balla, D.A. and Ciechetti, D.V. 1984. *Vineland Adaptive Behaviour Scale.* Circle Pines, MN: American Guidance Services.

Squire, J., Bricker, D. and Twombly, K. H 2012. The ages and stages questionnaire: Social-emotional: A validation study of a mother report questionnaire on a clinical mother-infant sample.

Stage, C. 2003. Classical test theory or item response theory: The Swedish experience. *A Journal of Psychological Methods* 2:57-78.

Stalman, B.F. and Pearson, P.D. 1991. Formal measures of early literacy. In L.M. Morrow and J.K. Smith Eds. *Assessment for instruction in early literacy* pp. 7-44. Englewood Cliffs, NJ: Prentice Hall.

Steinberg, L. and Thissie, D. 1996. Uses of item response theory and the test let concept in the measurement of psychopathology, *Psychological Methods*, 1:81-97.

Stout, W., Habing, B., Douglas, J., Kim, H. R., Roussos, L. and Zhang, J. 1996. Conditional covariance based non-parametric multidimensionality assessment. *Applied Psychological Measurements* 20.4:331-354.

Street, M. and Suite, N. W. 2000. *Early headstat program strategies socialization for infants and toddlers in the early headstat*. U. S. Department of Health and Human Services. Headstat National Resources Centre, Washington, D. C.

Strickland, D.S. and Schikedanz, J.A. 2004. Learning about print in preschool: working with letters, words, and beginning links with phonemic awareness. Newark, DE: International Reading Association.

Swaminathan . H. and Roggers -H.J 1990 . Detecting Differential item functioning using logistic Regression . Journal of educational measurement 27.4:361-370

Swaminathan, H. 1985. *Item response theory: principles and applications.* Boston, MA: Kluwer-Nijoff Publishing. Washington, DC: council of chief state school officers

Tabor, S.Q., Snow, C.E. and Dickinson, D.K. 2001. Homes and schools together: supporting language and literacy development. *Beginning literacy with language: young children learning at home and school* (pp. 313-334). Baltimore, MD Paul H. Brookes Publishing Co.

Tamivaara, M. Enright, S. 1986. On eliciting information: dialogues with child informants. *Anthropology and Education Quarterly* 17:218-238.

Tanner, D. and Tanner, L.N. 1975. *Curriculum development theory and practice*. New York: Macmillan Publishing Co., Inc.

Teacher's Rating of Oral Language and Literacy (TROLL) 2001. *A research based tool.* CIERA Report. Centre for the Improvement of Early Reading Achievement, Ann Arbor, MI.

Tealle, F. 1988. Develop cognitively appropriate assessment of reading and writing in the early childhood classroom. *Elementary School Journal* 89:173-183.

_____ 1992. The promise and the challenge of informal assessment in early literacy. In Morrow, L.M. and Smith, J.K. Eds. *Assessment for instruction in early literacy* pp. 45-61. New Jersey: Prentice Hall.

The Indian Journal of Pediatrics May 2004, Volume 71, issue J, pp 387 – 396

Thiessen, V. and Blasiu, S.J. 2008. Mathematics performance and mathematics, learning strategies: cognitive competences and construct differentiation. *International Journal of Educational Research*, 4:76, 362-371.

Thissen, D. and Orlando, M. 2001. Item response theory for item scored in two categories. In Thissen and Wainer, H. Eds. *Test scoring* (pp. 8=73-140). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Thorndike, R.L., Hagen, E.P. and Sattler, J.M. 2006. *Stanford-Binet Intelligence Scale* 4th Edn. Chicago: Riverside.

Thorpe, GL and Favia, A 2012. Information analysis using item response theory methodology: an introduction to selected programme an applications psychology faculty scholarship paper 20 . Retrieved 15th June, 2016 from http://digitalcommons.library, umaine.edu/psyfacupu/20.

Tomkowick, J. T. and Wright, K.R. 2007. Investigation of the effect of test equation and scoring methods on item parameter estimates and learner ability scores. A paper presented at the annual conference of American Educational Research Association, Chicago, April 10.

Tracey, D.H. and Morrow, L.M. 2006. *Lenses on reading.* New York: Guilford Press.

Traub, R. 1997. Classical test theory in historical perspective. *Educational Measurement: issues and practice.* 16.4:8-14.

Tyler, R.W. 1949. *Basic principles of curriculum and instruction*. London: The University of Chicago Press.

Ughamadu, R. A., Onwegbu, D. C. and Osunde, A. 1991. *Measurement and evaluation in education.* Benin City: World of Books Publication.

Vale, C. D. 2006. Computerised item banking in Downing, S. D. and Hala dyna, T. M. (Eds.) *The handbook of test development.* Routledge.

Velutino, B.C., Scanlon, D.M., and Sipay, E. 1997. Toward distinguishing between cognitive and experiential deficits as primary sources of difficulty in learning to read: the importance of early intervention in diagnostic specific reading disability. In B. Blachman Ed. *Foundation of reading acquisition and dyslexia: implications for early intervention* (pp. 347-380). Mahwah, NJ: Lawrence Erlbaum.

Vidya, T. 2014. Parental involvement in children education. *Journal of the Association for Childhood Education International* 10.3. Washington, DC.

Wagel, D. J., Marlin S. S., and Benneth, K.K. (2006). Contribution of the home literacy environment to pre-school aged children's emerging literacy and language skills in Early Child Development and Care. 176: 357- 378.

Wanda, B.H. 2013. Empowering learner independent reading. *Journal of the Association for Childhood Education International* 90.2. Washington, DC.

Wechsler, D. 2002. *Wechsler preschool and primary scale of intelligence (WPPSI-IU)*. (3rd Edn) San Antonio, TX: Psychological Corp.

Weis, B.C. 1999. Introductory statistics. USA: Addison Wesley Longman Inc.

Welberg, M. 2007. Measuring and detecting differential item functioning in criterion-referenced licensing test: a theoretical comparison of methods. Educational measurement No. 60 ISSN 1103-2685.

Wen-Wei, L., Rong-guey, H., Yung-chin, Y., Hsu-chen, C. 2012. Social behaviour and personality. *Academic Journal* 40.10:1679-1688.

Wesson, K.A. 2001. The 'volvo effect" – questioning standardised tests. *Young Children*, 56.2:16-18.

White, J. C. 2007. Learner-centred teacher-student relationship are effective: A meta-analysis. *Review of Education Research.*

Whitehouse. 2002. *Federal programmes for young children: Review and recommendations* (Vol. 13). Washington D. C.: U.S. Government Printing Office.

Whitmore, K. and Godman, P. 1998. Inside the whole language classroom. *School Administrator* 49.5:20-26.

Wiberg, M. 2004. Classical test theory vs. Item response theory: an assessment of the theory test in the Swedish driving – license test. *Educational measurement* No. 50 ISSN 1103-2685.

Williams, J.H. 2005. Cross-National variation in rural Mathematics Performance: A descriptive overview, *Journal of Research in Rural Education*, 20:1-8.

Wilson, R. 1980. Criterion-referenced testing. *Test Service Notebook* 37. New York: Psychological Corp.

Witt, J.C., Elliot, S.N., Dalylll, E.J., Gresham, F.M. and Kramer, J.J. 1998. *Assessment of at-risk and special needs children* (2nd Edn.) Boston: McGraw-Hill.

Worthan, S.C. 1998. Introduction in S.C. Wortham A handbook for preschool and elementary education (pp. 7-13). Olney, MD: Association for Childhood Education International.

_____ 2008. *Assessment in early childhood education*. 5th Edn. Upper Saddle River NJ: Prentice Hall.

Worthan, S. C., Barbour, A. and Desjean-Perrota, D. 1998. Introduction. *Portfolio assessment: a handbook for pre-school and elementary educators*. S.C. Wortham, A. Barbour and D. Desjean-Perrotta Eds. Olney, MD: Association for Childhood Education International.

Yen, W.M. 2002. Item response theory. Encylopedia of Educational Research 6th Ed. New York: Macmillan pp. 657-667.

Zanon, C., Hutz, C.S., Yoo, H. and Hambleton, R.K. 2016. An application of item response theory to psychological test development. *Psicologis: Reflexao e Critica* 29.18:1-10.

Zigler, E., and Valentine, J. (eds) (1979). Project Head start. A legacy of the war on poverty; New York. Free press.

Zimmerman, I.L., Steiner, V.G., and Pond, R.E. 2002. *Preschool Language Scale (PLS-4)*.4th Edn. San Antonio, TX: Psychological Corp.

Zumbo, B. D. and Hubley, A. 2017. Three generations of DIF analyses: considering where it has been, where it is now and where it is going. *Lawrence Assessment Quarterly* 4. 2: 223-233.

APPENDIX I

INTERNATIONAL CENTRE FOR EDUCATIONAL EVALUATION
UNIVERSITY OF IBADAN

ERLT Scoring Sheet

Centre Name: ……..

Centre Location:……..

Child's Name:…….. Age: ……

Child's Gender: ……Class: ……

SECTION B

Tick Yes for correct - carries 1

Tick No for missed or unanswered - which carries 0

Letters Identification

(i) The child is able to read capital letters

1 A		2 B		3 C		4 D		5 E		6 F	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
7 G		8 H		9 I		10 J		11 K		12 L	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
13 M		14 N		15 O		16 P		17 Q		18 R	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
19 S		20 T		21 U		22 V		23 W		24 X	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
25 Y		26 Z		Total = 26 Child's Total Score =							
Yes	No	Yes	No								

II The child is able to read juggled capital letters.

1/27 B		2 / 28 X		3 / 29 R		4 / 30 A		5 / 31 G		6 / 32 V	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
7 / 33 L		8 / 34 Z		9 / 35 N		10 / 36 Y		11 / 37 P		12 / 38 T	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
13/ 39 K		14 / 40 U		15 / 41 E		16 / 42 C		17 / 43 O		18 / 44 M	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Total = 18 Child's Total Score =											

III The child is able to identify small letters of the alphabets.

1 / 45 a		2 / 46 b		3 / 47 C		4 / 48 D		5 / 49 E		6 / 50 F	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
7 / 51 g		8 / 52 h		9 / 53 I		10 / 54 J		11 / 55 k		12 / 56 L	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
13 / 57 m		14 / 58 n		15 / 59 O		16 / 60 P		17 / 61 q		18 / 62 R	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No

19 / 63 s		20 / 64 t		21 / 65 U		22 / 66 V		23 / 67 w		24 / 68 X	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
25 / 69 y		26 / 70 z		Total = 26 Child's Total Score =							
Yes	No	Yes	No								

IV The child is able to identify juggled small letters of the alphabet.

2 / 71 c		3 / 72 m		4 / 73 p		5 / 74 V		6 / 75 n		7 / 76 Z	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
8 /77 u		9 / 78 a		10 / 79 G		11 / 80 R		12 / 81 s		13 / 82 o	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Total = 12 Child's Total Score =											

SECTION C

Identification

The child is able to identify objects, animals, birds , parts of human body and the rest

1 / 83 Pawpaw		2 / 84 Car		3 / 85 Tree		4 / 86 Ant		5 / 87 Pencil		6 / 88 Table	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
7 / 89 Spoon		8 / 90 Orange		9 / 91 Calendar		10 / 92 Ruler		11 / 93 Door		12 / 94 Mango	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No

13 / 95 Fan		14 / 96 Basket		15 / 97 Tyre		16 / 98 Flower		17 / 99 Shoe		18 / 100 Bicycle	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
19 / 101 House		20 / 102 Leg		21 / 103 Bird		22 / 104 Bell		23 / 105 Banana		24 / 106 Dog	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
25 / 107 Elephant		26 / 108 Iron		27 / 109 Umbrella		28 / 110 Rat		29 / 111 Kettle		30 / 112 Cap	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
31 / 113 Fork		32 / 114 Eye		33 / 115 Broom		34 / 116 Scissors		35 / 117 Ladder		36 / 118 Fingers	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
37 / 119 Egg		38 / 120 Bed		39 / 121 Pot		40 / 122 Radio		41 / 123 Baby		42 / 124 Car	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
43 / 125 Lion		44 / 126 Aeroplane		45 / 127 Money/5Naira		Total = 45 Child's Total Score =					
Yes	No	Yes	No	Yes	No						

SECTION D

Identification of colours.

The child is able to identify primary colours.

1 / 128 Blue		2 / 129 Red		Total = 2 Child's Total Score =
Yes	No	Yes	No	

SECTION E

Recognition of signs and symbols.

1 / 130 Stop		2 / 131 Ready		3 / 132 Go		Total = 3 Child's Total Score =
Yes	No	Yes	No	Yes	No	

SECTION F

READING FLUENCY

I. The child is able to read short sentences of common objects and animals.

1 / 133 Fish		2 / 134 Pot		3 / 135 Chair		4 / 136 Orange		5 / 137 Bottle		6 / 138 Window		7 / 139 Ball	
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
8 / 140 Lion		9 / 141 Jug		10 / 142 Bag		4/143 Nose		Total = 11 Child's Total Score =					
Yes	No	Yes	No	Yes	No	Yes	No						

SECTION G

RHYMES

I. The child is able to sing "One two buckle my Shoes"

1 / 144		2 / 145		3 / 146		4 / 147		5 / 148		Total = 5
One, two buckle my shoe		Three, four knock at the door		Five, six pick up the sticks		Seven, eight lay them straight		Nine, ten a big fat hen		Child's Total Score =
Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	

SECTION H

I. The child is able to read the activities in picture.

1 / 149		2 / 150		3 / 151		4 / 152		5 / 153		6 / 154		7 / 155	
The girl is picking mangoes/fruits		The boy is feeding or playing with the cat		The girl is washing plates		Two hens are standing/ hens are eating		The dog is eating bone/ The dog is putting something in its mouth		The boy is fetching water/ carrying pot		The woman/ mummy is bathing the baby	
Yes	No	Yes	No	Yes	Yes	Yes	No	Yes	No	Yes	No	Yes	No
8 / 156		9 / 157		Total = 9									
The woman/ mummy is washing cloths		The girl is sweeping the floor		Child's Total Score =									
Yes	No	Yes	No										

II. The child is able to read the picture (school environment).

1 / 158	2 / 159	3 / 160	4 / 161	5 / 162	6 / 163
The boys are	The boy is going to the	The girls are playing	The boy is sliding	The boys are playing	The girls are

running/ the boys are bending		class		with skipping rope/ are jumping				football.		clapping/ playing ten-ten/ dancing	
Yes	No	Yes	No	Yes	Yes	Yes	No	Yes	No	Yes	No
Total = 6 Child's Total Score =											

APPENDIX II

INTERNATIONAL CENTRE FOR EDUCATIONAL EVALUATION
UNIVERSITY OF IBADAN

MANUAL ON ADMINISTRATION OF EARLY READING LITERACY TEST (ERLT)

Early Reading Literacy Test (ERLT) was developed to test the major pre-reading skills of children that attend pre-schools in Nigeria as spelt out in the Nigeria Early Learning Development Standards and one-year compulsory pre-school education curriculum compiled by Nigerian Educational Research and Development Council (NERDC). ERLT is meant for children at the final pre-school classes. The test was culturally appropriate to Nigerian children. ERLT is in two sections: Section A generated demographic information such as age, gender and name of pre-school child. Information on the centre the child attends like the type of school and the school location were also included.

Section B – H consist of **226 test** items on: Letter identification, identification of objects, part of human body, fruits, animals, identification of colours, recognition of signs and symbols, ability to sing rhyme, reading fluency and picture reading. ERLT should be administered by: experienced pre-primary class teachers; Early Childhood care and Development Education specialists; Educational Evaluators; Child Psychologists and research assistants that have been trained on the purpose and the administration of the test.

The test administrator should ensure that testing security and integrity is maintained. There should be no leakage of the test. No reproduction of the test in any form should be allowed. Since it would be copyrighted The test administrators need to familiarise themselves with the children because children accept and relate freely with persons they are familiar with, not strangers. They should relate with the children, be part of the activities in the centres before the day of the test's administration. This calls for a pre-visit to the centres for external administrator, test administrator should also be familiar with the test, the materials and as well master the procedures before the testing day. Testing should be scheduled at a time that facilitates test takers' maximum performance. Preferably in the morning, before 12.00 noon. Test takers should be allowed to rest in between the test sections if need be, considering the children's attention span. No time limit should be set, every child must be allowed to

move at his/her own pace. However, the test administrator should note the average time spent on the testing. The sitting arrangement should be in a way that one will not distract another. The centre/classroom should remain the normal way it used to be, no need to remove displayed charts, toys and the rest. There should not be distraction around and within the venue. Each section of the test carries an instruction. The test administrator should give the instruction in his/her natural tone. The instruction may be repeated, if necessary. The test must be administered through one-on-one interaction with the test takers. It may be necessary to employ the service of more test administrators who should have been trained depending on the number of test takers. The test administrator should interact with a child at a time, and read out the instruction and allow each child to attempt all question. The test should be administered orally and the administrator should record the score immediately in the score sheet.

Section A

Demographic information

At this section,whoever wants to use this if necessary, will generate the demographic information of the children. These include child's

Name _____ _____ **Age _____**

Gender _____ **Type of school _____**

Location of school _____

Section B

LETTERS IDENTIFICATION (I – IV)

(i) This will be administered on one - on-one basis during which time the researcher/research assistant focuses a child, points to or touches an alphabet and asks the child to mention/identify the letters of alphabets one after the other;

(iii) No time limit is set, each child is allowed to move at his or her pace;

(iv) The researcher/research assistant ticks as the child responds.

Section C

IDENTIFICATION

(i) This will be done on a researcher/research assistant to a test taker interaction during which the researcher/assistant will do the following;
(ii) Points to or touches the object/animal/bird/human being one after the other and asks the child 'What is this?'
(iii) Allows each child to move at his/her pace;
(iv) Records only the child's first response;
(v) Record immediately as the child responds.

Section D

Identification of Colours

(i) This will also be administered one-on -one to a test taker by aresearcher/research assistant doing the following;
(ii) Touches the colour one after the other and asks the child "What colour is this?"
(iii) Allows each child to move at his/her pace for example. If it is Red that thechild identifies and points to, while the students is pointing to blue. recordsresponse against the appropriate item;
(iv) Allows the test taker to move at his/her pace.

Section E

Recognition of Signs and Symbols

(i) This section will be administered on the basis of a researcher/research assistant to a test taker at a time while the researcher/assistant do the following;
(ii) Starts singing "when you see a traffic light, there is something you must know";
(iii) While the test taker sings along and completes the song;
(iv) Researcher scores each line of the song as the child sings ;
(v) Records immediately in the score sheet;
(vi) Points to/touches the Nigeria flag, asks the child "What is this?"

Section F

Reading Fluency

(i) The administration of this section will be done on the basis of a researcher/research assistant to a test taker while the researcher/research assistant:
(ii) Leads them in reading the first sentence and thereafter asks the child to read the sentence one after the other;
(iii) Allows the child to attempt reading all the sentences;
(iv) Ensures the test take response is strictly in English language
(v) Ensures the child completes reading a sentence correctly before he/she is scored'1'

Section G

RHYME

(i) This section will be an interaction between a test taker and a researcher/research assistant at a time while the researcher/research assistant :Starts to sing the rhyme 'One, two buckle my shoe';

(ii) The test taker sings along and completes the remaining lines of the rhymes;

(iii) Records '1' on the score sheet, when the test taker sings a line of the rhymes correctly;

(iv) Ensures the child's response is strictly in English Language;

Section H

(a) Picture Reading (Home Environment)

(i) Section H will be administered on a test taker by a researcher/research assistant at a time while the researcher/research assistant;

(ii) Asks the test taker to read activities he/she sees in the picture;

(iii) Encourages the test taker to read out all the activities carried out in the picture by asking him/her "What other things can you see?";

(iv) Records '1 ' for the test taker as long as he/she has an idea of each scene.

(v) Ensures the test taker response is in English language.

(b) **Picture Reading (School Play ground)**

- Researcher/research assistant will follow the same procedure in the administration of this segment as in I (a)

APPENDIX III

ERLT REVIEW AT A-GLANCE

N A M E O F T E S T :	EARLY READING LITERACY TEST (ERLT)
A U T H O R	
P U B L I S H E R / Y E A R	
C L A S S R A N G E	CHILDREN IN THE FINAL PRE-SCHOOL LEVEL
N O R M I N G S A M P L E	CLASS, GENDER, AGE AND CENTRE TYPE
TOTAL NUMBER OF SAMPLE	7 7 6
L O C A T I O N	O Y O S T A T E , N I G E R I A
TIME REQUIREMENTS AND TESTING PACE	N O F I X E D T I M E Avarage of 15minutes per child

Standardisation

Age equivalent [ ] Scores [ ]

Percentile [ ]

Standard Scores [ ]

Stainne [ ]

DIFFERENTIAL ITEM FUNCTIONING

APPENDIX IV

Item parameter estimates in the DRA_ERLT using IRT framework

ITEM NUMBER	*IRT*	
	a	***b***
ERLT 1	0.61	-2.20
ERLT 2	1.17	-1.23
103	1.53	-2.50
ERLT 4	2.20	-1.84
ERLT 5	2.96	-1.62
ERLT 6	2.85	-1.42
ERLT 7	3.25	-1.07
ERLT 8	3.34	-1.00
ERLT 9	4.30	-0.95
ERLT 10	3.12	-0.88
ERLT 11	5.73	-0.90
ERLT 12	5.36	-0.83
ERLT 13	5.26	-0.80
ERLT 14	6.25	-0.80
ERLT 15	5.57	-0.84
ERLT 16	5.12	-0.72
ERLT 17	4.67	-0.60
ERLT 18	5.41	-0.63
ERLT 19	5.05	-0.69
ERLT 20	5.42	-0.64
ERLT 21	6.36	-0.59
ERLT 22	5.78	-3.57
ERLT 23	0.60	-0.59
ERLT 24	2.01	-5.71
ERLT 25	2.12	-0.85
ERLT 26	3.38	-0.87
ERLT 27	1.49	-2.37
ERLT 28	2.09	-4.19

ERLT 29	4.46	-0.51
ERLT 30	1.76	-3.65
ERLT 31	4.41	-0.62
ERLT 32	1.09	-2.78
ERLT 33	2.35	-0.78
ERLT 34	3.16	-5.83
ERLT 35	3.35	-6.75
ERLT 36	4.04	-3.70
ERLT 37	5.87	-0.49
ERLT 38	3.90	-0.71
ERLT 39	4.19	-0.72
ERLT 40	4.99	-0.62
ERLT 41	4.61	-0.59
ERLT 42	4.57	-0.65
ERLT 43	0.71	-4.28
ERLT 44	2.30	-3.53
ERLT 45	2.84	-4.65
ERLT 46	5.98	-0.51
ERLT 47	0.56	-0.77
ERLT 48	4.97	-0.47
ERLT 49	2.07	-1.39
ERLT 50	1.77	-1.60
ERLT 51	3.51	-0.79
ERLT 52	3.24	-0.72
ERLT 53	0.65	-1.66
ERLT 54	0.92	-1.09
ERLT 55	1.06	-2.72
ERLT 56	1.73	-1.85
ERLT 57	2.21	-1.55
ERLT 58	1.66	-1.43
ERLT 59	2.55	-0.99
ERLT 60	2.85	-0.94

ERLT 61	2.71	-0.82
ERLT 62	3.50	-0.79
ERLT 63	4.91	-0.87
ERLT 64	3.23	-0.65
ERLT 65	4.69	-0.75
ERLT 66	5.35	-0.76
ERLT 67	5.35	-0.81
ERLT 68	5.21	-0.70
ERLT 69	0.72	-0.22
ERLT 70	2.28	-0.59
ERLT 71	2.90	-0.67
ERLT 72	4.39	-0.61
ERLT 73	4.98	-0.55
ERLT 74	5.52	-0.54
ERLT 75	3.86	-0.58
ERLT 76	0.63	-0.52
ERLT 77	1.50	-0.77
ERLT 78	3.47	-0.69
ERLT 79	2.95	-4.67
ERLT 80	2.41	-3.53
ERLT 81	2.60	-1.92
ERLT 82	3.44	-3.77
ERLT 83	1.50	-3.64
ERLT 84	1.36	-4.52
ERLT 85	3.01	-0.67
ERLT 86	2.95	-0.64
ERLT 87	4.22	-0.84
ERLT 88	3.95	-5.01
ERLT 89	2.92	-3.45
ERLT 90	0.31	-4.55
ERLT 91	0.21	-0.46
ERLT 92	3.04	-3.80

ERLT 93	5.78	-0.61
ERLT 94	2.38	-0.44
ERLT 95	4.63	-0.51
ERLT 96	3.93	-5.57
ERLT 97	5.20	-3.15
ERLT 98	5.30	-4.52
ERLT 99	0.75	-1.32
ERLT 100	2.36	-0.62
ERLT 101	1.34	-3.46
ERLT 102	3.27	-0.56
ERLT 103	0.98	-0.71
ERLT 104	4.37	-0.81
ERLT 105	4.67	-0.24
ERLT 106	4.25	-0.61
ERLT 107	1.59	-0.49
ERLT 108	3.37	-0.41
ERLT 109	2.94	-2.05
ERLT 110	2.55	-3.62
ERLT 111	0.50	-1.86
ERLT 112	0.63	-0.45
ERLT 113	0.21	0.60
ERLT 114	1.69	-2.26
ERLT 115	1.90	-0.86
ERLT 116	1.09	-3.42
ERLT 117	1.93	-0.30
ERLT 118	1.80	-0.53
ERLT 119	1.68	-7.92
ERLT 120	2.04	-0.79
ERLT 121	0.12	-1.43
ERLT 122	1.62	-0.55
ERLT 123	0.71	-0.65
ERLT 124	2.28	-0.84

ERLT 125	3.20	-0.60
ERLT 126	2.25	-0.68
ERLT 127	3.07	-0.61
ERLT 128	3.55	-0.38
ERLT 129	2.88	-3.27
ERLT 130	2.18	-0.62
ERLT 131	2.12	-1.35
ERLT 132	2.35	0.15
ERLT 133	1.46	-0.73
ERLT 134	1.46	-0.90
ERLT 135	1.76	-0.66
ERLT 136	1.74	-0.79
ERLT 137	2.46	-0.58
ERLT 138	2.22	-0.41
ERLT 139	2.79	-0.07
ERLT 140	2.76	-4.47
ERLT 141	1.29	-3.57
ERLT 142	1.06	-0.55
ERLT 143	2.42	-0.36
ERLT 144	1.79	0.19
ERLT 145	3.02	-4.12
ERLT 146	1.72	0.11
ERLT 147	1.73	-3.22
ERLT 148	3.07	-5.02
ERLT 149	2.06	-1.70
ERLT 150	0.96	-1.64
ERLT 151	1.13	-1.22
ERLT 152	1.85	-1.98
ERLT 153	1.25	-1.50
ERLT 154	1.60	-0.81
ERLT 155	2.47	0.05
ERLT 156	2.68	0.14

ERLT 157	2.33	-3.19
ERLT 158	3.02	-3.40
ERLT 159	0.36	-5.77
ERLT 160	2.18	-0.93
ERLT 161	1.94	-0.36
ERLT 162	1.14	-3.74
ERLT 163	1.46	-3.09
ERLT 164	0.94	-3.29
ERLT 165	1.83	-6.89
ERLT 166	0.32	-3.48
ERLT 167	0.95	-0.34
ERLT 168	1.08	0.29
ERLT 169	0.41	-3.16
ERLT 170	1.20	-3.46
ERLT 171	2.72	0.46
ERLT 172	3.43	0.45
ERLT 173	3.28	-0.35
ERLT 174	2.28	-3.56
ERLT 175	2.46	-0.42
ERLT 176	2.81	-3.27
ERLT 177	1.80	-5.41
ERLT 178	2.59	-3.47
ERLT 179	2.59	-4.44
ERLT 180	2.67	0.58
ERLT 181	1.63	-3.18
ERLT 182	2.01	0.25
ERLT 183	2.00	-3.68
ERLT 184	1.58	-4.01
ERLT 185	1.89	-0.45
ERLT 186	2.54	-3.22
ERLT 187	3.18	-0.36
ERLT 188	3.96	0.61

ERLT 189	1.01	-4.06
ERLT 190	2.04	-3.33
ERLT 191	3.16	-4.18
ERLT 192	3.07	0.22
ERLT 193	3.48	-0.02
ERLT 194	2.71	-3.32
ERLT 195	5.13	0.85
ERLT 196	0.51	-4.03
ERLT 197	3.26	-0.25
ERLT 198	1.27	-0.24
ERLT 199	1.24	-0.22
ERLT 200	1.20	-0.21
ERLT 201	1.26	-0.25
ERLT 202	1.26	-0.73
ERLT 203	1.12	-4.53
ERLT 204	1.10	-3.32
ERLT 205	0.54	-3.52
ERLT 206	1.08	-3.37
ERLT 207	1.68	-5.58
ERLT 208	1.40	0.84
ERLT 209	0.54	-0.85
ERLT 210	0.62	0.49
ERLT 211	3.43	0.50
ERLT 212	3.96	-0.92
ERLT 213	0.80	0.39
ERLT 214	2.41	0.65
ERLT 215	0.80	0.00
ERLT 216	1.13	0.38
ERLT 217	1.89	-3.40
ERLT 218	1.90	-1.05
ERLT 219	0.88	-5.69
ERLT 220	1.17	-3.84

ERLT 221	1.03	0.46
ERLT 222	2.81	0.73
ERLT 223	2.49	-4.51
ERLT 224	3.05	0.53
ERLT 225	1.35	0.10
ERLT 226	1.12	0.77

How many of the items of the DRA-ERLT survived using IRT framework?

Item parameter estimates and retained Items

ITEM NUMBER	IRT				Decision
	a	Remark	b	Remark	
ERLT 1	0.61	Low	-2.20	Very Easy	Retain
ERLT 2	1.17	Moderate	-1.23	Easy	Retain
ERLT 3	1.53	High	-2.50	Very Easy	Retain
ERLT 4	2.20	Very High	-1.84	Very Easy	Retain
ERLT 5	2.96	Very High	-1.62	Easy	Retain
ERLT 6	2.85	Very High	-1.42	Easy	Retain
ERLT 7	3.25	Very High	-1.07	Easy	Retain
ERLT 8	3.34	Very High	-1.00	Easy	Retain
ERLT 9	4.30	Very High	-0.95	Easy	Retain
ERLT 10	3.12	Very High	-0.88	Easy	Retain
ERLT 11	5.73	Very High	-0.90	Easy	Retain
ERLT 12	5.36	Very High	-0.83	Easy	Retain
ERLT 13	5.26	Very High	-0.80	Easy	Retain
ERLT 14	6.25	Very High	-0.80	Easy	Retain
ERLT 15	5.57	Very High	-0.84	Easy	Retain
ERLT 16	5.12	Very High	-0.72	Easy	Retain
ERLT 17	4.67	Very High	-0.60	Easy	Retain
ERLT 18	5.41	Very High	-0.63	Easy	Retain
ERLT 19	5.05	Very High	-0.69	Easy	Retain
ERLT 20	5.42	Very High	-0.64	Easy	Retain
ERLT 21	6.36	Very High	-0.59	Moderate	Retain
ERLT 22	5.78	Very High	-3.57	Very Easy	Retain

ERLT 23	0.60	Low	-0.59	Very easy	Retain
ERLT 24	2.01	Very High	-5.71	Very Easy	Retain
ERLT 25	2.12	Very High	-0.85	Easy	Retain
ERLT 26	3.38	Very High	-0.87	Easy	Retain
RERLT 27	1.49	High	-2.37	Very easy	Retain
ERLT 28	2.09	Very High	-4.19	Very easy	Reject
ERLT 29	4.46	Very High	-0.51	Moderate	Retain
ERLT 30	1.76	Very High	-3.65	Very easy	Reject
ERLT 31	4.41	Very High	-0.62	Easy	Retain
ERLT 32	1.09	Moderate	-2.78	Very easy	Retain
ERLT 33	2.35	Very High	-0.78	Easy	Retain
ERLT 34	3.16	Very High	-5.83	Very easy	Reject
ERLT 35	3.35	Very High	-6.75	Very easy	Reject
ERLT 36	4.04	Very High	-3.70	Very easy	Reject
ERLT 37	5.87	Very High	-0.49	Moderate	Retain
ERLT 38	3.90	Very High	-0.71	Easy	Retain
ERLT 39	4.19	Very High	-0.72	Easy	Retain
ERLT 40	4.99	Very High	-0.62	Easy	Retain
ERLT 41	4.61	Very High	-0.59	Moderate	Retain
ERLT 42	4.57	Very High	-0.65	Easy	Retain
ERLT 43	0.71	Moderate	-4.28	Very easy	Reject
ERLT 44	2.30	Very High	-3.53	Very easy	Reject
ERLT 45	2.84	Very High	-4.65	Very easy	Reject
ERLT 46	5.98	Very High	-0.51	Moderate	Retain
ERLT 47	0.56	Low	-0.77	Easy	Retain
ERLT 48	4.97	Very High	-0.47	Moderate	Retain
ERLT 49	2.07	Very High	-1.39	Easy	Retain
ERLT 50	1.77	Very High	-1.60	Easy	Retain
ERLT 51	3.51	Very High	-0.79	Easy	Retain
ERLT 52	3.24	Very High	-0.72	Easy	Retain
ERLT 53	0.65	Moderate	-1.66	Easy	Retain
iERLT 54	0.92	Moderate	-1.09	Easy	Retain

ERLT 55	1.06	Moderate	-2.72	Very easy	Retain
ERLT 56	1.73	Very high	-1.85	Very easy	Retain
ERLT 57	2.21	Very high	-1.55	Very easy	Retain
ERLT 58	1.66	High	-1.43	Easy	Retain
ERLT 59	2.55	Very high	-0.99	Easy	Retain
ERLT 60	2.85	Very high	-0.94	Easy	Retain
ERLT 61	2.71	Very high	-0.82	Easy	Retain
ERLT 62	3.50	Very high	-0.79	Easy	Retain
ERLT 63	4.91	Very high	-0.87	Easy	Retain
ERLT 64	3.23	Very high	-0.65	Easy	Retain
ERLT 65	4.69	Very high	-0.75	Easy	Retain
ERLT 66	5.35	Very high	-0.76	Easy	Retain
ERLT 67	5.35	Very high	-0.81	Easy	Retain
ERLT 68	5.21	Very high	-0.70	Easy	Retain
ERLT 69	0.72	Moderate	-0.22	Moderate	Retain
ERLT 70	2.28	Very	-0.59	Moderate	Retain
ERLT 71	2.90	Very high	-0.67	Easy	Retain
ERLT 72	4.39	Very high	-0.61	Easy	Retain
ERLT 73	4.98	Very high	-0.55	Moderate	Retain
ERLT 74	5.52	Very high	-0.54	Moderate	Retain
ERLT 75	3.86	Very high	-0.58	Moderate	Retain
ERLT 76	0.63	Low	-0.52	Moderate	Retain
ERLT 77	1.50	High	-0.77	Easy	Retain
ERLT 78	3.47	Very high	-0.69	Easy	Retain
ERLT 79	2.95	Very high	-4.67	Very easy	Reject
ERLT 80	2.41	Very high	-3.53	Very easy	Reject
ERLT 81	2.60	Very high	-1.92	Very easy	Retain
ERLT 82	3.44	Very high	-3.77	Very easy	Reject
ERLT 83	1.50	High	-3.64	Very easy	Reject
ERLT 84	1.36	High	-4.52	Very easy	Reject
ERLT 85	3.01	Very high	-0.67	Easy	Retain
ERLT 86	2.95	Very high	-0.64	Easy	Retain

ERLT 87	4.22	Very high	-0.84	Easy	Retain
ERLT 88	3.95	Very high	-5.01	Very easy	Reject
ERLT 89	2.92	Very high	-3.45	Very easy	Reject
ERLT 90	0.31	Very high	-4.55	Very easy	Reject
ERLT 91	0.21	Very low	-0.46	Moderate	Retain
ERLT 92	3.04	Very high	-3.80	Very easy	Reject
ERLT 93	5.78	Very high	-0.61	Easy	Retain
ERLT 94	2.38	Very high	-0.44	Moderate	Retain
ERLT 95	4.63	Very high	-0.51	Moderate	Retain
ERLT 96	3.93	Very high	-5.57	Very easy	Reject
ERLT 97	5.20	Very high	-3.15	Very easy	Reject
ERLT 98	5.30	Very high	-4.52	Very esy	Reject
ERLT 99	0.75	Moderate	-1.32	Easy	Retain
ERLT 100	2.36	Very high	-0.62	Easy	Retain
ERLT 101	1.34	Moderate	-3.46	Very easy	Reject
ERLT 102	3.27	Very high	-0.56	Moderate	Retain
ERLT 103	0.98	Moderate	-0.71	Easy	Retain
ERLT 104	4.37	Very high	-0.81	Easy	Retain
ERLT 105	4.67	Very high	-0.24	Moderate	Retain
ERLT 106	4.25	Very high	-0.61	Easy	Retain
ERLT 107	1.59	High	-0.49	Moderate	Retain
ERLT 108	3.37	Very high	-0.41	Moderate	Retain
ERLT 109	2.94	Very high	-2.05	Very easy	Retain
ERLT 110	2.55	Very high	-3.62	Very easy	Reject
ERLT 111	0.50	Low	-1.86	Very easy	Retain
ERLT 112	0.63	Low	-0.45	Moderate	Retain
ERLT 113	0.21	Very low	0.60	Moderate	Retain
ERLT 114	1.69	High	-2.26	Very easy	Retain
ERLT 115	1.90	Very high	-0.86	Easy	Retain
ERLT 116	1.09	Moderate	-3.42	Very easy	Reject
ERLT 117	1.93	Very high	-0.30	Moderate	Retain
ERLT 118	1.80	Very high	-0.53	Moderate	Retain

ERLT 119	1.68	High	-7.92	Very easy	Reject
ERLT 120	2.04	Very high	-0.79	Easy	Retain
ERLT 121	0.12	Very low	-1.43	Easy	Retain
ERLT 122	1.62	High	-0.55	Moderate	Retain
ERLT 123	0.71	Moderate	-0.65	Easy	Retain
ERLT 124	2.28	Very high	-0.84	Easy	Retain
ERLT 125	3.20	Very high	-0.60	Easy	Retain
ERLT 126	2.25	Very high	-0.68	Easy	Retain
ERLT 127	3.07	Very high	-0.61	Easy	Retain
ERLT 128	3.55	Very high	-0.38	Moderate	Retain
ERLT 129	2.88	Very high	-3.27	Very easy	Reject
ERLT 130	2.18	Very high	-0.62	Easy	Retain
ERLT 131	2.12	Very high	-1.35	Easy	Retain
ERLT 132	2.35	Very high	0.15	Moderate	Retain
ERLT 133	1.46	High	-0.73	Easy	Retain
ERLT 134	1.46	High	-0.90	Easy	Retain
ERLT 135	1.76	Very high	-0.66	Easy	Retain
ERLT 136	1.74	Very high	-0.79	Easy	Retain
ERLT 137	2.46	Very high	-0.58	Moderate	Retain
ERLT 138	2.22	Very high	-0.41	Moderate	Retain
ERLT 139	2.79	Very high	-0.07	Moderate	Retain
ERLT 140	2.76	Very high	-4.47	Very easy	Reject
ERLT 141	1.29	Moderate	-3.57	Very easy	Reject
ERLT 142	1.06	Moderate	-0.55	Moderate	Retain
ERLT 143	2.42	Very high	-0.36	Moderate	Retain
ERLT 144	1.79	Very high	0.19	Moderate	Retain
ERLT 145	3.02	Very high	-4.12	Very easy	Reject
ERLT 146	1.72	Very high	0.11	Moderate	Retain
ERLT 147	1.73	Very high	-3.22	Very easy	Reject
ERLT 148	3.07	Very high	-5.02	Very easy	Reject
ERLT 149	2.06	Very high	-1.70	Easy	Retain
ERLT 150	0.96	Moderate	-1.64	Easy	Retain

ERLT 151	1.13	Moderate	-1.22	Easy	Retain
ERLT 152	1.85	Very high	-1.98	Very easy	Retain
ERLT 153	1.25	Moderate	-1.50	Easy	Retain
ERLT 154	1.60	High	-0.81	Easy	Retain
ERLT 155	2.47	Very high	0.05	Moderate	Retain
ERLT 156	2.68	Very high	0.14	Moderate	Retain
ERLT 157	2.33	Very high	-3.19	Very easy	Reject
ERLT 158	3.02	Very high	-3.40	Very easy	Reject
ERLT 159	0.36	Low	-5.77	Very easy	Reject
ERLT 160	2.18	Very high	-0.93	Easy	Retain
ERLT 161	1.94	Very high	-0.36	Moderate	Retain
ERLT 162	1.14	Moderate	-3.74	Very easy	Reject
ERLT 163	1.46	High	-3.09	Very easy	Reject
ERLT 164	0.94	Moderate	-3.29	Very easy	Reject
ERLT 165	1.83	Very high	-6.89	Very easy	Reject
ERLT 166	0.32	Very low	-3.48	Very easy	Reject
ERLT 167	0.95	Moderate	-0.34	Moderate	Retain
ERLT 168	1.08	Moderate	0.29	Moderate	Retain
ERLT 169	0.41	Moderate	-3.16	Very easy	Reject
ERLT 170	1.20	Moderate	-3.46	Very easy	Reject
ERLT 171	2.72	Very high	0.46	Moderate	Retain
ERLT 172	3.43	Very high	0.45	Moderate	Retain
ERLT 173	3.28	Very high	-0.35	Moderate	Retain
ERLT 174	2.28	Very high	-3.56	Very easy	Reject
ERLT 175	2.46	Very high	-0.42	Moderate	Retain
ERLT 176	2.81	Very high	-3.27	Very easy	Reject
ERLT 177	1.80	Very high	-5.41	Very easy	Reject
ERLT 178	2.59	Very high	-3.47	Very easy	Reject
ERLT 179	2.59	Very high	-4.44	Very easy	Reject
ERLT 180	2.67	Very high	0.58	Moderate	Retain
ERLT 181	1.63	High	-3.18	Very easy	Reject
ERLT 182	2.01	Very high	0.25	Moderate	Retain

ERLT 183	2.00	Very high	-3.68	Very easy	Reject
ERLT 184	1.58	High	-4.01	Very easy	Reject
ERLT 185	1.89	Very high	-0.45	Moderate	Retain
ERLT 186	2.54	Very high	-3.22	Very easy	Reject
ERLT 187	3.18	Very high	-0.36	Moderate	Retain
ERLT 188	3.96	Very high	0.61	Hard	Retain
ERLT 189	1.01	Moderate	-4.06	Very easy	Reject
ERLT 190	2.04	Very high	-3.33	Very easy	Reject
ERLT 191	3.16	Very high	-4.18	Very easy	Reject
ERLT 192	3.07	Very high	0.22	Moderate	Reject
ERLT 193	3.48	Very high	-0.02	Moderate	Retain
ERLT 194	2.71	Very high	-3.32	Very easy	Reject
ERLT 195	5.13	Very high	0.85	Hard	Retain
ERLT 196	0.51	Low	-4.03	Very easy	Reject
ERLT 197	3.26	Very high	-0.25	Moderate	Retain
ERLT 198	1.27	Moderate	-0.24	Moderate	Retain
ERLT 199	1.24	Moderate	-0.22	Moderate	Retain
ERLT 200	1.20	Moderate	-0.21	Moderate	Retain
ERLT 201	1.26	Moderate	-0.25	Moderate	Retain
ERLT 202	1.26	Moderate	-0.73	Easy	Retain
ERLT 203	1.12	Moderate	-4.53	Very easy	Reject
ERLT 204	1.10	Moderate	-3.32	Very easy	Reject
ERLT 205	0.54	Low	-3.52	Very easy	Reject
ERLT 206	1.08	Moderate	-3.37	Very easy	Reject
ERLT 207	1.68	High	-5.58	Very easy	Reject
ERLT 208	1.40	High	0.84	Hard	Retain
ERLT 209	0.54	Low	-0.85	Easy	Retain
ERLT 210	0.62	Low	0.49	Moderate	Retain
ERLT 211	3.43	Very high	0.50	Moderate	Retain
ERLT 212	3.96	Very high	-0.92	Easy	Retain
ERLT 213	0.80	Moderate	0.39	Moderate	Retain
ERLT 214	2.41	Very high	0.65	Hard	Retain

ERLT 215	0.80	Moderate	0.00	Moderate	Retain
ERLT 216	1.13	Moderate	0.38	Moderate	Retain
ERLT 217	1.89	Very high	-3.40	Very easy	Reject
ERLT 218	1.90	Very high	-1.05	Easy	Retain
ERLT 219	0.88	Moderate	-5.69	Very easy	Reject
ERLT 220	1.17	Moderate	-3.84	Very easy	Reject
ERLT 221	1.03	Moderate	0.46	Moderate	Retain
ERLT 222	2.81	Very high	0.73	Hard	Retain
ERLT 223	2.49	Very high	-4.51	Very easy	Reject
ERLT 224	3.05	Very high	0.53	Moderate	Retain
ERLT 225	1.35	High	0.10	Moderate	Retain
ERLT 226	1.12	Moderate	0.77	Hard	Retain

Analysis of DIF with respect to Gender

Item number	M-H CHI	M-H LOR	LOR SE	LOR Z	BD	ETS	Remarks
Item 1	0.00	0.00	0.00	0.00	0.00	NO DIF	
Item 2	0.01	-0.60	0.99	-0.60	1.54	NO DIF	
Item 3	0.00	-0.35	0.81	-0.43	0.34	NO DIF	
Item 4	0.05	0.39	0.69	0.56	0.06	NO DIF	
Item 5	0.03	0.08	0.66	0.12	1.41	NO DIF	
Item 6	0.82	-0.57	0.51	-1.12	2.61	NO DIF	
Item 7	0.00	0.09	0.46	0.19	0.49	NO DIF	
Item 8	0.68	-0.44	0.42	-1.04	0.38	NO DIF	
Item 9	0.82	-0.54	0.47	-1.14	2.84	NO DIF	
Item 10	0.02	0.03	0.44	0.07	0.43	NO DIF	
Item 11	0.18	-0.36	0.52	-0.69	1.12	NO DIF	
Item 12	0.02	-0.25	0.57	-0.43	0.32	NO DIF	
Item 13	0.00	-0.21	0.65	-0.33	0.33	NO DIF	
Item 14	0.11	0.45	0.69	0.65	1.01	NO DIF	
Item 15	0.00	0.16	0.58	0.27	0.04	NO DIF	
Item 16	0.03	0.07	0.57	0.12	2.51	NO DIF	

Item 17	0.01	0.12	0.57	0.20	0.65	NO DIF	
Item 18	0.02	-0.09	0.60	-0.15	0.15	NO DIF	
Item 19	0.22	-0.33	0.48	-0.69	5.60	NO DIF	
Item 20	3.50	-1.05	0.55	-1.89	15.23	NO DIF	
Item 21	0.17	-0.40	0.57	-0.71	0.12	NO DIF	
Item 22	0.83	-0.69	0.59	-1.17	1.31	NO DIF	
Item 23	0.71	0.30	0.30	0.98	7.34	NO DIF	
Item 24	5.51	-0.97	0.41	-2.39	3.76	DIF	Male examinees
Item 25	3.42	-0.71	0.36	-1.97	0.09	NO DIF	
Item 26	2.66	-0.77	0.42	-1.81	1.14	NO DIF	
Item 27	1.21	-1.24	0.82	-1.51	4.62	NO DIF	
Item 28	0.41	-0.30	0.37	-0.83	0.13	NO DIF	
Item 29	0.01	0.06	0.45	0.13	0.05	NO DIF	
Item 30	0.01	-0.08	0.49	-0.17	2.51	NO DIF	
Item 31	0.66	-0.54	0.51	-1.05	0.59	NO DIF	
Item 32	0.20	-0.50	0.66	-0.76	4.28	NO DIF	
Item 33	0.01	-0.02	0.34	-0.07	0.69	NO DIF	
Item 34	0.78	-0.39	0.38	-1.05	1.09	NO DIF	
Item 35	0.20	-0.24	0.38	-0.63	1.36	NO DIF	
Item 36	0.24	-0.31	0.44	-0.69	1.18	NO DIF	
Item 37	0.06	-0.30	0.55	-0.54	0.34	NO DIF	
Item 38	0.05	-0.19	0.44	-0.42	1.87	NO DIF	
Item 39	4.93	-1.50	0.62	2.42	1.93	DIF	Favour female examinees
Item 40	0.34	-0.44	0.53	-0.83	6.54	NO DIF	
Item 41	0.92	-0.76	0.58	-1.31	0.08	NO DIF	
Item 42	0.04	0.02	0.49	0.05	0.29	NO DIF	
Item 43	1.75	0.48	0.32	1.49	1.46	NO DIF	
Item 44	13.57	-2.07	0.64	-3.24	2.24	DIF	Male examinees
Item 45	1.01	-0.55	0.46	-1.21	2.45	NO DIF	
Item 46	1.76	-0.85	0.54	-1.59	5.80	NO DIF	
Item 47	0.00	0.04	0.27	0.16	2.67	NO DIF	

Item 48	0.41	-0.49	0.51	-0.95	0.00	NO DIF	
Item 49	0.06	0.20	0.42	0.47	0.31	NO DIF	
Item 50	0.07	-0.20	0.42	-0.47	2.83	NO DIF	
Item 51	1.38	-0.71	0.50	-1.42	4.15	NO DIF	
Item 52	0.02	-0.15	0.43	-0.35	0.12	NO DIF	
Item 53	1.36	-1.03	0.71	-1.45	1.02	NO DIF	
Item 54	0.03	-0.14	0.75	-0.19	0.00	NO DIF	
Item 55	0.16	-0.55	0.71	-0.78	0.23	NO DIF	
Item 56	0.00	-0.13	0.54	-0.25	0.19	NO DIF	
Item 57	0.03	0.07	0.58	0.11	0.06	NO DIF	
Item 58	0.21	-0.27	0.40	-0.66	0.40	NO DIF	
Item 59	0.01	-0.03	0.38	-0.08	0.06	NO DIF	
Item 60	0.12	-0.20	0.38	-0.53	0.06	NO DIF	
Item 61	2.09	-0.65	0.41	-1.59	1.20	NO DIF	
Item 62	3.37	-0.89	0.46	-1.94	1.96	NO DIF	
Item 63	1.93	-0.85	0.54	-1.58	0.33	NO DIF	
Item 64	3.06	-0.93	0.48	-1.92	0.92	NO DIF	
Item 65	0.05	-0.05	0.63	-0.08	0.41	NO DIF	
Item 66	0.02	0.27	0.61	0.45	0.00	NO DIF	
Item 67	0.04	0.08	0.63	0.13	0.06	NO DIF	
Item 68	0.35	0.64	0.70	0.92	0.08	NO DIF	
Item 69	0.81	0.31	0.31	1.02	2.94	NO DIF	
Item 70	3.98	-0.85	0.42	-2.02	10.87	DIF	Male examinees
Item 71	1.78	-0.71	0.47	-1.52	1.51	NO DIF	
Item 72	0.21	0.37	0.52	0.71	0.00	NO DIF	
Item 73	0.02	-0.24	0.57	-0.42	0.62	NO DIF	
Item 74	0.03	0.27	0.58	0.46	0.77	NO DIF	
Item 75	0.12	-0.27	0.46	-0.59	0.29	NO DIF	
Item 76	0.42	0.21	0.27	0.77	3.09	NO DIF	
Item 77	0.20	-0.29	0.45	-0.64	0.05	NO DIF	
Item 78	3.20	-0.72	0.39	-1.83	1.69	NO DIF	
Item 79	0.79	0.42	0.38	1.10	0.02	NO DIF	

Item 80	1.48	-0.64	0.45	-1.43	0.41	NO DIF	
Item 81	1.08	-0.79	0.58	-1.35	0.43	NO DIF	
Item 82	0.32	0.29	0.40	0.74	0.07	NO DIF	
Item 83	1.62	-0.62	0.43	-1.46	0.68	NO DIF	
Item 84	4.21	-1.08	0.49	-2.23	5.67	DIF	Male examinees
Item 85	0.01	-0.18	0.49	-0.36	0.26	NO DIF	
Item 86	0.03	0.23	0.54	0.44	0.28	NO DIF	
Item 87	1.38	-0.69	0.49	-1.40	0.56	NO DIF	
Item 88	1.71	0.33	0.24	1.39	0.72	NO DIF	
Item 89	1.19	0.29	0.24	1.21	0.53	NO DIF	
Item 90	0.18	-0.24	0.39	-0.61	0.82	NO DIF	
Item 91	0.44	-0.50	0.53	-0.94	0.50	NO DIF	
Item 92	0.05	0.14	0.35	0.40	0.00	NO DIF	
Item 93	0.86	-0.58	0.50	-1.16	0.16	NO DIF	
Item 94	2.10	-0.97	0.55	-1.75	0.53	NO DIF	
Item 95	0.04	0.03	0.48	0.06	0.98	NO DIF	
Item 96	0.00	-0.13	0.50	-0.27	2.74	NO DIF	
Item 97	0.00	0.04	0.29	0.13	9.93	NO DIF	
Item 98	3.21	-0.76	0.40	-1.89	9.15	NO DIF	
Item 99	3.22	-0.64	0.33	-1.91	0.34	NO DIF	
Item 100	0.16	-0.24	0.41	-0.58	5.36	NO DIF	
Item 101	0.79	-0.40	0.37	-1.08	0.11	NO DIF	
Item 102	0.48	-0.46	0.50	-0.92	0.15	NO DIF	
Item 103	2.05	-0.88	0.55	-1.60	7.98	NO DIF	
Item 104	0.05	-0.26	0.52	-0.49	0.01	NO DIF	
Item 105	0.28	0.19	0.28	0.68	1.55	NO DIF	
Item 106	0.87	0.57	0.47	1.20	0.07	NO DIF	
Item 107	7.70	-1.19	0.44	-2.72	4.86	DIF	Male examinees
Item 108	0.16	0.21	0.35	0.59	0.00	NO DIF	
Item 109	0.08	0.49	0.73	0.68	0.09	NO DIF	
Item 110	0.43	0.19	0.24	0.79	0.10	NO DIF	
Item 111	1.16	0.34	0.28	1.20	0.86	NO DIF	

Item 112	0.01	0.01	0.31	0.04	0.29	NO DIF	
Item 113	1.21	0.35	0.28	1.24	0.57	NO DIF	
Item 114	2.07	0.78	0.45	1.73	0.00	NO DIF	
Item 115	1.79	-0.67	0.45	-1.51	0.75	NO DIF	
Item 116	0.22	-0.21	0.33	-0.64	2.66	NO DIF	
Item 117	1.38	0.41	0.30	1.36	0.21	NO DIF	
Item 118	5.68	0.68	0.28	2.43	0.21	DIF	Female examinees
Item 119	3.18	0.66	0.33	1.98	0.62	NO DIF	
Item 120	2.07	0.43	0.29	1.52	2.99	NO DIF	
Item 121	2.65	-0.57	0.34	-1.68	12.72	NO DIF	
Item 122	0.00	0.02	0.24	0.07	0.08	NO DIF	
Item 123	0.22	0.22	0.34	0.65	0.11	NO DIF	
Item 124	0.03	0.20	0.47	0.43	0.09	NO DIF	
Item 125	5.24	0.87	0.35	2.45	3.97	DIF	Female examinees
Item 126	0.76	0.47	0.43	1.08	1.67	NO DIF	
Item 127	0.00	0.15	0.50	0.31	0.10	NO DIF	
Item 128	1.68	0.59	0.39	1.51	3.60	NO DIF	
Item 129	1.49	0.51	0.36	1.44	0.31	NO DIF	
Item 130	0.08	0.13	0.30	0.43	0.41	NO DIF	
Item 131	0.75	0.38	0.37	1.02	0.95	NO DIF	
Item 132	4.07	0.70	0.33	2.12	1.93	DIF	Female examinees
Item 133	3.74	0.78	0.37	2.12	5.35	DIF	Female examinees
Item 134	0.38	0.21	0.28	0.75	1.40	NO DIF	
Item 135	0.88	0.35	0.32	1.11	1.00	NO DIF	
Item 136	0.01	0.02	0.35	0.06	0.97	NO DIF	
Item 137	0.01	-0.03	0.36	-0.07	0.90	NO DIF	
Item 138	0.98	0.44	0.37	1.19	1.81	NO DIF	
Item 139	0.14	-0.19	0.35	-0.54	0.08	NO DIF	

Item 140	1.27	0.37	0.29	1.28	1.52	NO DIF	
Item 141	2.75	0.61	0.33	1.86	0.06	NO DIF	
Item 142	4.93	1.05	0.42	2.52	0.46	DIF	Female examinees
Item 143	0.80	0.31	0.30	1.06	3.13	NO DIF	
Item 144	1.90	0.58	0.38	1.54	0.77	NO DIF	
Item 145	2.43	0.52	0.31	1.70	0.03	NO DIF	
Item 146	5.03	0.78	0.33	2.36	1.47	DIF	Female examinees
Item 147	1.14	0.57	0.45	1.27	0.21	NO DIF	
Item 148	0.02	0.05	0.48	0.11	0.54	NO DIF	
Item 149	7.90	1.10	0.38	2.90	3.79	DIF	Favour female examinees
Item 150	0.01	0.03	0.33	0.08	0.09	NO DIF	
Item 151	0.00	-0.05	0.34	-0.14	0.18	NO DIF	
Item 152	0.00	-0.03	0.33	-0.10	2.93	NO DIF	
Item 153	0.19	-0.32	0.48	-0.68	0.02	NO DIF	
Item 154	0.02	0.11	0.35	0.31	1.30	NO DIF	
Item 155	0.01	-0.10	0.36	-0.28	0.19	NO DIF	
Item 156	1.91	0.61	0.39	1.57	2.41	NO DIF	
Item 157	2.02	0.69	0.44	1.57	1.10	NO DIF	
Item 158	0.02	0.06	0.53	0.12	0.07	NO DIF	
Item 159	1.61	0.31	0.22	1.37	0.72	NO DIF	
Item 160	0.44	0.30	0.36	0.84	0.45	NO DIF	
Item 161	2.57	0.59	0.33	1.79	1.00	NO DIF	
Item 162	0.02	0.06	0.25	0.25	0.00	NO DIF	
Item 163	0.77	-0.39	0.37	-1.06	0.08	NO DIF	
Item 164	0.52	0.24	0.28	0.86	0.01	NO DIF	
Item 165	1.16	-0.41	0.34	-1.22	0.33	NO DIF	
Item 166	0.01	-0.05	0.22	-0.21	1.71	NO DIF	
Item 167	0.01	-0.05	0.23	-0.21	2.98	NO DIF	
Item 168	0.34	0.16	0.24	0.70	1.70	NO DIF	

Item 169	0.00	-0.02	0.20	-0.08	0.77	NO DIF	
Item 170	2.64	-0.46	0.26	-1.77	7.03	NO DIF	
Item 171	1.43	-0.55	0.40	-1.37	1.39	NO DIF	
Item 172	4.97	-1.38	0.63	-2.17	56.00	DIF	Favour male examinees
Item 173	9.24	-1.87	0.67	-2.77	19.79	DIF	Favour male examinees
Item 174	2.47	-0.63	0.36	-1.75	3.65	NO DIF	
Item 175	1.65	-0.61	0.41	-1.49	0.91	NO DIF	
Item 176	0.03	-0.17	0.45	-0.39	0.01	NO DIF	
Item 177	0.00	-0.05	0.36	-0.15	0.31	NO DIF	
Item 178	0.02	-0.12	0.38	-0.33	0.00	NO DIF	
Item 179	0.08	-0.22	0.44	-0.50	0.34	NO DIF	
Item 180	0.00	-0.10	0.42	-0.24	0.00	NO DIF	
Item 181	0.07	-0.25	0.48	-0.51	0.00	NO DIF	
Item 182	0.02	-0.01	0.33	-0.02	0.06	NO DIF	
Item 183	0.08	-0.18	0.38	-0.47	0.00	NO DIF	
Item 184	0.39	0.49	0.54	0.90	1.45	NO DIF	
Item 185	0.44	0.25	0.31	0.81	0.28	NO DIF	
Item 186	0.34	0.32	0.41	0.79	1.18	NO DIF	
Item 187	0.07	0.16	0.36	0.43	3.44	NO DIF	
Item 188	0.60	-0.52	0.52	-1.01	2.15	NO DIF	
Item 189	1.71	0.36	0.26	1.43	11.86	NO DIF	
Item 190	5.66	-0.82	0.33	-2.45	18.90	DIF	Favour male examinees
Item 191	0.52	-0.45	0.48	-0.93	2.04	NO DIF	
Item 192	1.97	-0.54	0.35	-1.54	23.18	NO DIF	
Item 193	0.05	-0.22	0.48	-0.46	2.17	NO DIF	
Item 194	0.04	0.12	0.33	0.37	0.14	NO DIF	
Item 195	0.01	-0.08	0.54	-0.16	0.48	NO DIF	
Item 196	0.99	0.28	0.25	1.13	0.01	NO DIF	
Item 197	1.79	0.51	0.34	1.50	3.82	NO DIF	

Item 198	1.71	0.35	0.25	1.41	0.06	NO DIF	
Item 199	1.31	0.30	0.24	1.24	0.05	NO DIF	
Item 200	1.59	0.32	0.24	1.35	0.09	NO DIF	
Item 201	1.13	0.28	0.24	1.16	0.05	NO DIF	
Item 202	1.27	0.30	0.24	1.24	0.12	NO DIF	
Item 203	0.12	-0.11	0.24	-0.46	1.52	NO DIF	
Item 204	0.47	-0.19	0.24	-0.80	3.63	NO DIF	
Item 205	0.75	-0.21	0.22	-0.97	2.90	NO DIF	
Item 206	0.12	-0.11	0.23	-0.47	1.48	NO DIF	
Item 207	0.03	0.00	0.37	0.00	0.00	NO DIF	
Item 208	0.00	0.07	0.31	0.22	0.53	NO DIF	
Item 209	0.06	-0.09	0.24	-0.37	0.42	NO DIF	
Item 210	0.00	-0.02	0.23	-0.09	0.21	NO DIF	
Item 211	0.71	-0.75	0.62	-1.20	0.55	NO DIF	
Item 212	0.03	-0.40	0.73	-0.56	0.37	NO DIF	
Item 213	1.04	0.26	0.23	1.12	0.02	NO DIF	
Item 214	0.00	0.09	0.45	0.21	0.02	NO DIF	
Item 215	1.52	0.34	0.25	1.35	3.35	NO DIF	
Item 216	2.21	-0.38	0.24	-1.58	3.47	NO DIF	
Item 217	0.02	0.01	0.33	0.04	0.01	NO DIF	
Item 218	0.28	0.28	0.38	0.73	0.17	NO DIF	
Item 219	0.08	-0.11	0.26	-0.42	0.02	NO DIF	
Item 220	0.02	0.01	0.30	0.02	0.02	NO DIF	
Item 221	1.21	0.32	0.26	1.24	0.01	NO DIF	
Item 222	0.33	0.45	0.54	0.82	0.01	NO DIF	
Item 223	0.55	-0.53	0.51	-1.04	0.48	NO DIF	
Item 224	0.01	0.17	0.51	0.33	1.96	NO DIF	
Item 225	0.00	0.02	0.27	0.08	0.01	NO DIF	
Item 226	0.31	-0.25	0.34	-0.75	0.54	NO DIF	

Analysis of DIF with respect to School Type

Item number	MH CHI	MH LOR	LOR SE	LOR Z	BD	ETS	Remarks
Item 1	0.00	0.00	0.00	0.00	0.00	NO DIF	
Item 2	0.37	-0.05	1.05	-0.05	3.72	NO DIF	
Item 3	3.05	0.00	0.00	0.00	0.00	NO DIF	
Item 4	4.13	-1.59	0.76	-2.11	0.39	DIF	Favour private examinees
Item 5	1.88	-0.92	0.59	-1.55	0.27	NO DIF	
Item 6	0.98	-0.54	0.47	-1.16	1.38	NO DIF	
Item 7	0.82	-0.47	0.43	-1.09	1.19	NO DIF	
Item 8	2.69	-0.65	0.39	-1.65	0.01	NO DIF	
Item 9	2.44	-0.85	0.50	-1.72	11.90	NO DIF	
Item 10	2.11	-0.89	0.54	-1.65	10.07	NO DIF	
Item 11	1.09	0.69	0.55	1.26	6.19	NO DIF	
Item 12	0.08	0.02	0.57	0.03	0.08	NO DIF	
Item 13	0.99	0.61	0.53	1.15	6.07	NO DIF	
Item 14	0.04	-0.25	0.56	-0.46	7.64	NO DIF	
Item 15	1.14	0.70	0.58	1.21	5.98	NO DIF	
Item 16	0.49	0.50	0.55	0.91	6.86	NO DIF	
Item 17	0.00	-0.10	0.48	-0.20	2.31	NO DIF	
Item 18	0.00	-0.15	0.54	-0.28	0.18	NO DIF	
Item 19	2.06	0.83	0.52	1.60	10.37	NO DIF	
Item 20	0.32	0.41	0.52	0.78	6.95	NO DIF	
Item 21	3.53	-1.25	0.64	-1.94	21.06	NO DIF	
Item 22	1.73	-0.71	0.53	-1.35	1.10	NO DIF	
Item 23	7.97	-1.23	0.45	-2.70	0.40	DIF	Favour private examinees
Item 24	10.81	-1.49	0.47	-3.21	0.75	DIF	Favour private examinees

Item 25	0.09	-0.18	0.39	-0.47	0.06	NO DIF	
Item 26	1.27	-0.56	0.44	-1.27	0.93	NO DIF	
Item 27	0.20	0.47	0.63	0.75	10.75	NO DIF	
Item 28	1.71	0.49	0.34	1.44	7.96	NO DIF	
Item 29	1.63	-0.54	0.41	-1.31	0.06	NO DIF	
Item 30	0.05	-0.19	0.44	-0.43	1.73	NO DIF	
Item 31	5.04	1.07	0.50	2.12	9.50	DIF	Favour public examinees
Item 32	0.03	-0.17	0.78	-0.21	0.09	NO DIF	
Item 33	0.10	0.15	0.32	0.46	0.30	NO DIF	
Item 34	0.02	0.01	0.34	0.02	0.13	NO DIF	
Item 35	0.89	0.40	0.37	1.09	0.46	NO DIF	
Item 36	0.00	0.12	0.44	0.28	0.60	NO DIF	
Item 37	3.25	-0.84	0.50	-1.67	7.05	NO DIF	
Item 38	0.03	0.18	0.46	0.39	0.11	NO DIF	
Item 39	0.01	0.07	0.47	0.15	2.83	NO DIF	
Item 40	0.01	0.09	0.52	0.17	0.27	NO DIF	
Item 41	0.39	0.37	0.48	0.78	5.05	NO DIF	
Item 42	0.36	0.35	0.45	0.76	6.92	NO DIF	
Item 43	1.24	-0.59	0.47	-1.27	0.41	NO DIF	
Item 44	0.01	-0.13	0.44	-0.29	0.19	NO DIF	
Item 45	0.71	0.56	0.52	1.08	0.01	NO DIF	
Item 46	0.11	-0.21	0.42	-0.51	0.16	NO DIF	
Item 47	1.36	0.52	0.41	1.27	0.01	NO DIF	
Item 48	0.49	-0.36	0.43	-0.84	0.09	NO DIF	
Item 49	0.93	0.40	0.37	1.08	0.07	NO DIF	
Item 50	0.56	0.39	0.41	0.94	0.54	NO DIF	
Item 51	1.77	0.67	0.45	1.49	0.09	NO DIF	
Item 52	7.52	1.10	0.42	2.60	9.67	DIF	Favour public examinees

Item 53	3.47	1.68	0.78	2.16	0.96	NO DIF	
Item 54	0.62	0.72	0.66	1.09	12.99	NO DIF	
Item 55	0.00	-0.31	0.76	-0.41	0.87	NO DIF	
Item 56	0.87	-0.57	0.51	-1.13	3.45	NO DIF	
Item 57	0.56	-0.49	0.51	-0.96	1.71	NO DIF	
Item 58	1.44	-0.50	0.39	-1.29	0.01	NO DIF	
Item 59	0.99	-0.43	0.38	-1.13	0.11	NO DIF	
Item 60	3.42	-0.76	0.41	-1.87	0.29	NO DIF	
Item 61	0.25	0.26	0.39	0.67	4.34	NO DIF	
Item 62	0.04	0.01	0.49	0.02	0.20	NO DIF	
Item 63	0.00	0.19	0.63	0.30	1.58	NO DIF	
Item 64	1.03	0.45	0.40	1.11	13.63	NO DIF	
Item 65	1.91	1.14	0.74	1.55	0.93	NO DIF	
Item 66	1.45	1.36	0.92	1.48	0.18	NO DIF	
Item 67	3.72	2.38	1.22	1.95	1.56	NO DIF	
Item 68	1.51	1.17	0.79	1.48	15.78	NO DIF	
Item 69	0.03	0.02	0.42	0.05	0.07	NO DIF	
Item 70	0.00	0.14	0.47	0.29	0.24	NO DIF	
Item 71	0.16	0.37	0.54	0.68	0.00	NO DIF	
Item 72	0.01	-0.09	0.53	-0.17	1.13	NO DIF	
Item 73	0.56	-0.56	0.57	-0.98	0.05	NO DIF	
Item 74	1.10	-0.57	0.49	-1.15	3.43	NO DIF	
Item 75	0.02	-0.13	0.40	-0.32	5.12	NO DIF	
Item 76	3.21	-0.98	0.51	-1.93	0.09	NO DIF	
Item 77	0.22	0.39	0.54	0.73	0.60	NO DIF	
Item 78	2.54	0.67	0.41	1.64	7.92	NO DIF	
Item 79	1.58	-0.48	0.34	-1.38	0.14	NO DIF	
Item 80	0.54	-0.35	0.40	-0.87	0.01	NO DIF	
Item 81	0.00	-0.12	0.54	-0.22	5.55	NO DIF	
Item 82	0.41	0.31	0.39	0.79	18.72	NO DIF	
Item 83	4.62	1.08	0.49	2.20	3.37	DIF	Favour public

							examinees
Item 84	0.03	-0.17	0.45	-0.38	0.12	NO DIF	
Item 85	3.88	1.15	0.54	2.12	10.71	DIF	Favour public examinees
Item 86	3.68	1.17	0.63	1.86	12.58	NO DIF	
Item 87	3.05	1.18	0.62	1.91	0.21	NO DIF	
Item 88	0.05	0.12	0.33	0.37	0.05	NO DIF	
Item 89	0.06	0.13	0.33	0.40	0.08	NO DIF	
Item 90	9.80	1.23	0.41	2.96	16.40	DIF	Favour public examinees
Item 91	3.51	-0.92	0.50	-1.85	3.85	NO DIF	
Item 92	0.00	-0.06	0.33	-0.17	0.55	NO DIF	
Item 93	0.14	0.27	0.45	0.59	9.88	NO DIF	
Item 94	0.03	0.05	0.50	0.09	6.02	NO DIF	
Item 95	0.65	-0.44	0.46	-0.95	1.55	NO DIF	
Item 96	1.07	-0.62	0.49	-1.27	7.68	NO DIF	
Item 97	0.08	0.19	0.41	0.46	2.52	NO DIF	
Item 98	0.17	0.30	0.48	0.64	0.39	NO DIF	
Item 99	0.44	0.29	0.35	0.83	0.12	NO DIF	
Item 100	0.21	-0.27	0.42	-0.64	1.74	NO DIF	
Item 101	0.73	-0.40	0.40	-0.99	0.10	NO DIF	
Item 102	1.02	0.63	0.53	1.19	0.63	NO DIF	
Item 103	0.00	0.16	0.52	0.30	1.86	NO DIF	
Item 104	2.25	0.98	0.58	1.69	3.36	NO DIF	
Item 105	4.91	-0.67	0.31	-2.20	0.37	DIF	Favour private examinees
Item 106	10.29	-1.32	0.44	-3.02	0.67	DIF	Favour private examinees
Item 107	9.27	0.97	0.36	2.66	41.23	DIF	Favour

							public examinees
Item 108	8.45	0.92	0.36	2.58	36.15	DIF	Favour public examinees
Item 109	1.54	-1.64	1.19	-1.37	0.87	NO DIF	
Item 110	0.45	0.22	0.27	0.80	0.31	NO DIF	
Item 111	0.02	-0.24	0.56	-0.42	0.11	NO DIF	
Item 112	2.66	0.59	0.35	1.69	1.29	NO DIF	
Item 113	12.43	-1.29	0.38	-3.43	4.93	DIF	Favour private examinees
Item 114	3.00	-0.69	0.42	-1.66	3.99	NO DIF	
Item 115	5.20	1.37	0.51	2.69	0.69	DIF	Favour public examinees
Item 116	5.23	0.75	0.33	2.31	15.21	DIF	Favour public examinees
Item 117	0.24	-0.20	0.32	-0.62	0.16	NO DIF	
Item 118	1.51	-0.47	0.34	-1.37	3.61	NO DIF	
Item 119	13.58	1.14	0.33	3.40	19.85	DIF	Favour public examinees
Item 120	2.49	0.60	0.36	1.66	0.08	NO DIF	
Item 121	0.90	-0.43	0.40	-1.09	0.24	NO DIF	
Item 122	17.00	1.07	0.28	3.86	0.06	DIF	Favour public examinees
Item 123	20.24	1.58	0.39	4.08	35.49	DIF	Favour public examinees
Item 124	0.18	0.19	0.35	0.55	2.41	NO DIF	

Item 125	0.01	-0.08	0.32	-0.24	0.12	NO DIF	
Item 126	4.99	1.01	0.46	2.18	9.21	DIF	Favour public examinees
Item 127	1.12	0.52	0.44	1.19	1.29	NO DIF	
Item 128	0.26	0.24	0.36	0.68	2.03	NO DIF	
Item 129	1.11	-0.41	0.35	-1.18	3.03	NO DIF	
Item 130	0.00	-0.05	0.33	-0.15	3.81	NO DIF	
Item 131	3.62	-0.76	0.38	-2.03	4.91	NO DIF	
Item 132	2.80	0.65	0.35	1.88	1.84	NO DIF	
Item 133	13.83	-1.23	0.36	-3.42	21.49	DIF	Favour private examinees
Item 134	11.50	1.16	0.35	3.32	4.01	DIF	Favour public examinees
Item 135	0.00	-0.04	0.30	-0.14	2.08	NO DIF	
Item 136	0.38	0.26	0.36	0.74	2.07	NO DIF	
Item 137	6.87	1.11	0.43	2.57	0.13	DIF	Favour public examinees
Item 138	0.33	0.30	0.39	0.76	0.01	NO DIF	
Item 139	5.63	0.96	0.38	2.51	5.72	DIF	Favour public examinees
Item 140	8.92	-0.90	0.31	-2.88	12.19	DIF	Favour private examinees
Item 141	36.66	-2.25	0.49	-4.61	18.33	DIF	Favour private examinees
Item 142	6.56	-0.97	0.39	-2.50	3.98	DIF	Favour private

							examinees
Item 143	0.06	-0.11	0.30	-0.37	0.98	NO DIF	
Item 144	0.76	-0.38	0.39	-0.98	0.81	NO DIF	
Item 145	17.76	-1.55	0.39	-3.96	13.78	DIF	Favour private examinees
Item 146	34.16	-1.99	0.39	-5.06	49.10	DIF	Favour private examinees
Item 147	23.78	-2.63	0.57	-4.63	21.93	DIF	Favour private examinees
Item 148	13.69	-1.77	0.57	-3.09	69.93	DIF	Favour private examinees
Item 149	19.36	-1.54	0.40	-3.80	29.60	DIF	Favour private examinees
Item 150	0.82	-0.42	0.38	-1.10	1.97	NO DIF	
Item 151	0.24	0.22	0.34	0.66	7.63	NO DIF	
Item 152	3.01	0.70	0.36	1.93	0.06	NO DIF	
Item 153	0.09	0.23	0.44	0.51	0.40	NO DIF	
Item 154	0.33	-0.27	0.37	-0.73	0.40	NO DIF	
Item 155	5.29	-0.88	0.37	-2.37	0.70	DIF	Favour private examinees
Item 156	20.54	-1.92	0.54	-3.56	45.20	DIF	Favour private examinees
Item 157	28.99	-3.04	0.87	-3.49	15.88	DIF	Favour private examinees
Item 158	22.46	-3.03	0.78	-3.91	9.99	DIF	Favour

							private examinees
Item 159	5.54	-0.79	0.33	-2.40	0.60	DIF	Favour private examinees
Item 160	12.21	-1.21	0.39	-3.09	2.27	DIF	Favour private examinees
Item 161	16.11	-1.32	0.35	-3.74	8.42	DIF	Favour private examinees
Item 162	10.39	0.89	0.28	3.13	5.57	DIF	Favour public examinees
Item 163	0.00	-0.18	0.57	-0.32	0.48	NO DIF	
Item 164	19.09	1.32	0.32	4.14	12.40	DIF	Favour public examinees
Item 165	0.13	-0.20	0.37	-0.53	3.25	NO DIF	
Item 166	17.06	1.04	0.27	3.80	9.77	DIF	Favour public examinees
Item 167	8.72	0.76	0.27	2.81	7.69	DIF	Favour public examinees
Item 168	0.07	0.11	0.29	0.39	0.39	NO DIF	
Item 169	0.01	-0.06	0.25	-0.24	0.02	NO DIF	
Item 170	0.57	0.27	0.31	0.87	1.02	NO DIF	
Item 171	8.06	-1.31	0.57	-2.32	10.37	DIF	Favour private examinees
Item 172	2.37	-0.97	0.66	-1.48	11.91	NO DIF	
Item 173	2.57	-1.11	0.70	-1.59	3.19	NO DIF	

Item 174	13.29	1.15	0.38	3.05	17.27	DIF	Favour public examinees
Item 175	0.01	0.04	0.35	0.11	0.04	NO DIF	
Item 176	5.08	1.20	0.52	2.33	0.14	DIF	Favour public examinees
Item 177	0.06	0.15	0.37	0.40	5.90	NO DIF	
Item 178	14.55	1.37	0.43	3.18	8.84	DIF	Favour public examinees
Item 179	3.69	0.83	0.43	1.91	0.03	NO DIF	
Item 180	8.04	1.17	0.45	2.57	3.07	DIF	Favour public examinees
Item 181	4.52	1.20	0.57	2.12	0.28	DIF	Favour public examinees
Item 182	1.52	0.47	0.37	1.27	1.87	NO DIF	
Item 183	1.41	0.62	0.47	1.33	2.09	NO DIF	
Item 184	0.12	-0.20	0.40	-0.48	19.61	NO DIF	
Item 185	0.13	0.18	0.35	0.51	7.05	NO DIF	
Item 186	17.37	1.69	0.49	3.42	6.94	DIF	Favour public examinees
Item 187	4.04	-0.96	0.50	-1.92	70.78	DIF	Favour private examinees
Item 188	1.61	0.57	0.45	1.27	15.05	NO DIF	
Item 189	0.18	0.28	0.44	0.63	0.22	NO DIF	
Item 190	1.00	0.48	0.43	1.11	0.29	NO DIF	
Item 191	8.89	1.26	0.53	2.39	26.55	DIF	Favour public

							examinees
Item 192	2.49	-0.74	0.43	-1.72	10.68	NO DIF	
Item 193	11.69	-1.38	0.50	-2.76	40.07	DIF	Favour private examinees
Item 194	7.54	-1.04	0.41	-2.54	7.81	DIF	Favour private examinees
Item 195	2.57	0.76	0.49	1.55	19.64	NO DIF	
Item 196	1.65	0.45	0.36	1.24	1.88	NO DIF	
Item 197	7.25	1.33	0.51	2.60	55.38	DIF	Favour public examinees
Item 198	9.66	-0.99	0.31	-3.16	0.75	DIF	Favour private examinees
Item 199	9.04	-0.94	0.31	-3.06	0.78	DIF	Favour private examinees
Item 200	8.29	-0.91	0.31	-2.96	0.26	DIF	Favour private examinees
Item 201	7.19	-0.82	0.30	-2.76	0.19	DIF	Favour private examinees
Item 202	8.69	-0.90	0.30	-3.00	0.32	DIF	Favour private examinees
Item 203	0.95	-0.29	0.26	-1.10	0.08	NO DIF	
Item 204	1.00	-0.30	0.26	-1.13	0.04	NO DIF	
Item 205	0.58	-0.23	0.26	-0.90	0.14	NO DIF	
Item 206	0.82	-0.27	0.26	-1.03	0.11	NO DIF	
Item 207	0.00	-0.04	0.36	-0.12	0.00	NO DIF	

Item 208	3.42	-0.83	0.45	-1.83	5.26	NO DIF	
Item 209	15.28	1.24	0.34	3.60	0.02	DIF	Favour public examinees
Item 210	28.19	1.50	0.31	4.86	8.17	DIF	Favour public examinees
Item 211	4.22	-2.68	1.26	-2.13	4.17	DIF	Favour private examinees
Item 212	7.05	-2.84	1.20	-2.37	15.17	DIF	Favour private examinees
Item 213	23.33	1.19	0.27	4.40	14.73	DIF	Favour public examinees
Item 214	6.24	-0.97	0.46	-2.09	29.75	DIF	Favour private examinees
Item 215	0.01	-0.09	0.34	-0.26	2.17	NO DIF	
Item 216	1.93	0.43	0.30	1.43	0.46	NO DIF	
Item 217	2.39	-0.59	0.38	-1.55	5.11	NO DIF	
Item 218	0.00	0.06	0.39	0.15	0.00	NO DIF	
Item 219	39.12	1.55	0.29	5.38	24.88	DIF	Favour public examinees
Item 220	7.77	-1.10	0.44	-2.54	8.56	DIF	Favour private examinees
Item 221	10.39	0.92	0.29	3.22	5.49	DIF	Favour public examinees
Item 222	4.78	-1.21	0.59	-2.05	10.36	DIF	Favour

							private examinees
Item 223	5.17	-1.41	0.61	-2.31	3.13	DIF	Favour private examinees
Item 224	6.06	-1.03	0.57	-1.80	43.51	NO DIF	
Item 225	0.04	-0.11	0.31	-0.35	1.43	NO DIF	
Item 226	0.16	-0.20	0.38	-0.53	0.92	NO DIF	

Presents analysis of DIF with respect to school location

Item number	MH CHI	MH LOR	LOR SE	LOR Z	BD	ETS	Remarks
Item 1	0.00	0.00	0.00	0.00	0.00	NO DIF	
Item 2	2.44	0.00	0.00	0.00	0.00	NO DIF	
Item 3	7.98	0.00	0.00	0.00	0.00	NO DIF	
Item 4	2.85	-1.21	0.65	-1.86	0.41	NO DIF	
Item 5	0.85	-0.74	0.64	-1.15	0.26	NO DIF	
Item 6	0.98	-0.81	0.63	-1.29	0.00	NO DIF	
Item 7	9.94	-1.66	0.56	-2.95	4.25	DIF	Favour urban examinees
Item 8	11.47	-1.77	0.57	-3.10	1.76	DIF	Favour urban examinees
Item 9	3.85	-1.12	0.56	-2.02	3.66	DIF	Favour urban examinees
Item 10	4.86	-1.59	0.71	-2.24	0.53	DIF	Favour urban examinees
Item 11	4.08	-1.57	0.78	-2.01	1.28	DIF	Favour urban examinees
Item 12	8.48	0.00	0.00	0.00	0.00	NO DIF	
Item 13	2.91	-1.42	0.82	-1.74	0.83	NO DIF	
Item 14	4.70	-1.62	0.79	-2.05	6.42	DIF	Favour urban examinees
Item 15	0.01	0.28	0.65	0.43	0.00	NO DIF	
Item 16	0.00	0.27	0.70	0.39	0.05	NO DIF	
Item 17	1.01	-1.15	0.88	-1.30	0.04	NO DIF	
Item 18	0.02	-0.31	0.68	-0.46	0.06	NO DIF	
Item 19	1.77	0.84	0.58	1.46	1.38	NO DIF	
Item 20	4.83	1.48	0.72	2.06	7.13	DIF	Favour rural examinees
Item 21	0.74	0.58	0.56	1.02	16.70	NO DIF	
Item 22	0.98	0.72	0.62	1.16	4.34	NO DIF	
Item 23	9.81	-1.92	0.65	-2.93	0.56	DIF	Favour urban

							examinees
Item 24	1.99	0.45	0.33	1.36	44.83	NO DIF	
Item 25	5.40	0.78	0.37	2.13	15.80	DIF	Favour rural examinees
Item 26	0.02	-0.04	0.43	-0.09	0.01	NO DIF	
Item 27	0.65	-0.87	0.77	-1.13	0.44	NO DIF	
Item 28	1.73	0.58	0.39	1.48	0.67	NO DIF	
Item 29	4.41	1.11	0.50	2.22	20.35	DIF	Favour rural examinees
Item 30	0.23	0.34	0.48	0.70	6.63	NO DIF	
Item 31	5.31	1.27	0.53	2.39	6.37	DIF	Favour rural examinees
Item 32	0.06	0.55	0.84	0.66	1.32	NO DIF	
Item 33	2.07	0.57	0.37	1.54	3.62	NO DIF	
Item 34	0.02	-0.03	0.39	-0.07	1.80	NO DIF	
Item 35	5.51	0.96	0.42	2.29	10.67	DIF	Favour rural examinees
Item 36	7.99	1.36	0.51	2.66	14.73	DIF	Favour rural examinees
Item 37	15.62	3.13	1.09	2.87	12.68	DIF	Favour rural examinees
Item 38	0.05	0.20	0.47	0.43	3.96	NO DIF	
Item 39	3.79	1.20	0.60	2.00	11.92	NO DIF	
Item 40	6.70	1.53	0.60	2.54	13.17	DIF	Favour rural examinees
Item 41	6.44	2.26	1.10	2.05	1.71	DIF	Favour rural examinees
Item 42	8.62	2.09	0.85	2.45	5.27	DIF	Favour rural examinees
Item 43	2.36	0.66	0.38	1.75	2.75	NO DIF	
Item 44	31.04	2.15	0.56	3.87	42.86	DIF	Favour rural examinees
Item 45	38.22	0.00	0.00	0.00	0.00	NO DIF	

Item 46	17.63	2.84	1.05	2.69	20.95	DIF	Favour rural examinees
Item 47	11.94	-1.09	0.37	-2.93	14.66	DIF	Favour urban examinees
Item 48	30.50	0.00	0.00	0.00	0.00	NO DIF	
Item 49	8.90	1.31	0.46	2.89	1.94	DIF	Favour rural examinees
Item 50	2.19	0.71	0.44	1.61	0.82	NO DIF	
Item 51	5.35	1.15	0.52	2.23	0.03	DIF	Favour rural examinees
Item 52	9.55	1.63	0.58	2.84	0.97	DIF	Favour rural examinees
Item 53	0.01	0.26	0.80	0.32	0.83	NO DIF	
Item 54	0.01	-0.31	0.90	-0.35	0.01	NO DIF	
Item 55	0.29	-0.64	0.75	-0.85	0.04	NO DIF	
Item 56	0.31	-0.41	0.52	-0.78	0.09	NO DIF	
Item 57	0.65	-0.54	0.55	-0.99	0.15	NO DIF	
Item 58	0.66	-0.44	0.44	-1.01	0.09	NO DIF	
Item 59	11.93	-1.58	0.51	-3.13	3.50	DIF	Favour urban examinees
Item 60	8.37	-1.38	0.49	-2.81	1.90	DIF	Favour urban examinees
Item 61	8.15	-1.40	0.50	-2.82	5.44	DIF	Favour urban examinees
Item 62	6.05	-1.66	0.68	-2.46	3.84	DIF	Favour urban examinees
Item 63	1.18	-1.00	0.73	-1.37	0.70	NO DIF	
Item 64	0.70	-0.62	0.56	-1.11	0.01	NO DIF	
Item 65	0.04	-0.36	0.69	-0.52	0.74	NO DIF	
Item 66	0.96	-0.75	0.64	-1.16	8.29	NO DIF	
Item 67	0.57	0.69	0.69	1.00	0.02	NO DIF	
Item 68	0.84	-1.16	0.94	-1.23	0.13	NO DIF	
Item 69	6.71	-1.18	0.47	-2.51	1.37	DIF	Favour urban

							examinees
Item 70	14.36	1.25	0.42	2.97	50.50	DIF	Favour rural examinees
Item 71	24.02	2.39	0.66	3.61	9.67	DIF	Favour rural examinees
Item 72	1.81	0.90	0.62	1.47	1.38	NO DIF	
Item 73	2.72	1.70	0.90	1.90	2.39	NO DIF	
Item 74	0.08	0.35	0.61	0.58	1.66	NO DIF	
Item 75	0.27	-0.43	0.55	-0.78	0.15	NO DIF	
Item 76	31.11	-2.33	0.58	-4.03	13.47	DIF	Favour urban examinees
Item 77	0.40	0.43	0.51	0.84	0.08	NO DIF	
Item 78	0.00	-0.09	0.39	-0.23	1.64	NO DIF	
Item 79	2.38	-0.62	0.37	-1.67	1.33	NO DIF	
Item 80	0.59	-0.51	0.52	-0.98	0.20	NO DIF	
Item 81	5.42	-1.77	0.74	-2.39	0.05	DIF	Favour urban examinees
Item 82	1.62	0.60	0.42	1.42	0.05	NO DIF	
Item 83	14.37	2.07	0.67	3.12	8.81	DIF	Favour rural examinees
Item 84	15.81	1.95	0.59	3.33	19.51	DIF	Favour rural examinees
Item 85	0.46	0.44	0.49	0.88	0.07	NO DIF	
Item 86	0.01	0.30	0.65	0.46	0.09	NO DIF	
Item 87	0.43	0.58	0.60	0.96	0.03	NO DIF	
Item 88	0.07	0.09	0.26	0.37	19.26	NO DIF	
Item 89	1.02	-0.27	0.26	-1.01	15.31	NO DIF	
Item 90	2.42	0.84	0.48	1.74	2.11	NO DIF	
Item 91	6.75	1.84	0.69	2.66	5.09	DIF	Favour rural examinees
Item 92	6.26	0.93	0.37	2.52	3.69	DIF	Favour rural examinees
Item 93	3.57	1.34	0.61	2.19	2.03	NO DIF	

Item 94	0.67	0.76	0.70	1.09	2.72	NO DIF	
Item 95	4.41	1.51	0.65	2.33	4.03	DIF	Favour rural examinees
Item 96	5.82	1.63	0.66	2.45	8.76	DIF	Favour rural examinees
Item 97	0.91	0.48	0.40	1.19	2.22	NO DIF	
Item 98	51.39	3.90	1.07	3.63	8.06	DIF	Favour rural examinees
Item 99	18.09	1.28	0.34	3.76	13.25	DIF	Favour rural examinees
Item 100	4.03	0.90	0.47	1.91	11.65	DIF	Favour rural examinees
Item 101	0.53	-0.51	0.55	-0.94	0.02	NO DIF	
Item 102	6.91	1.55	0.62	2.51	4.82	DIF	Favour rural examinees
Item 103	0.37	0.46	0.56	0.82	0.07	NO DIF	
Item 104	0.12	0.29	0.52	0.57	0.48	NO DIF	
Item 105	13.61	-1.93	0.60	-3.20	4.88	DIF	Favour urban examinees
Item 106	0.05	0.21	0.48	0.44	0.00	NO DIF	
Item 107	0.42	0.41	0.49	0.84	0.38	NO DIF	
Item 108	0.81	0.53	0.49	1.07	0.19	NO DIF	
Item 109	0.00	-0.32	0.85	-0.37	0.00	NO DIF	
Item 110	16.56	-1.59	0.40	-3.96	0.87	DIF	Favour urban examinees
Item 111	41.51	-3.26	0.70	-4.64	1.52	DIF	Favour urban examinees
Item 112	9.90	-1.46	0.46	-3.14	0.86	DIF	Favour urban examinees
Item 113	8.10	-1.38	0.49	-2.85	2.61	DIF	Favour urban examinees
Item 114	0.99	-0.69	0.58	-1.20	1.47	NO DIF	
Item 115	0.16	0.33	0.52	0.63	0.01	NO DIF	

Item 116	0.94	-0.45	0.39	-1.15	0.17	NO DIF	
Item 117	7.30	-1.31	0.52	-2.51	1.28	DIF	Favour urban examinees
Item 118	10.32	-1.97	0.67	-2.93	2.62	DIF	Favour urban examinees
Item 119	2.27	-0.76	0.47	-1.62	1.79	NO DIF	
Item 120	10.90	-1.34	0.42	-3.20	1.46	DIF	Favour urban examinees
Item 121	19.72	1.36	0.38	3.56	30.05	DIF	Favour rural examinees
Item 122	0.42	0.20	0.27	0.73	2.37	NO DIF	
Item 123	1.14	-0.62	0.48	-1.29	0.36	NO DIF	
Item 124	1.60	0.80	0.53	1.49	0.25	NO DIF	
Item 125	7.15	-1.24	0.45	-2.75	0.09	DIF	Favour urban examinees
Item 126	0.34	0.44	0.53	0.84	0.82	NO DIF	
Item 127	0.02	0.07	0.53	0.14	0.01	NO DIF	
Item 128	0.05	0.00	0.46	-0.01	0.15	NO DIF	
Item 129	1.46	-0.88	0.59	-1.50	0.80	NO DIF	
Item 130	6.65	-1.68	0.67	-2.51	3.65	DIF	Favour urban examinees
Item 131	3.77	-1.26	0.63	-2.02	0.56	NO DIF	
Item 132	8.49	-1.19	0.41	-2.88	0.56	DIF	Favour urban examinees
Item 133	10.00	-1.29	0.45	-2.85	7.20	DIF	Favour urban examinees
Item 134	3.99	0.78	0.37	2.14	0.00	DIF	Favour rural examinees
Item 135	0.02	0.01	0.34	0.03	0.05	NO DIF	
Item 136	0.04	0.21	0.48	0.45	0.57	NO DIF	
Item 137	7.38	1.19	0.44	2.71	0.80	DIF	Favour rural examinees
Item 138	0.00	-0.12	0.47	-0.25	1.47	NO DIF	

Item 139	4.69	1.43	0.68	2.12	0.34	DIF	Favour rural examinees
Item 140	0.49	-0.35	0.41	-0.84	2.32	NO DIF	
Item 141	1.09	-0.51	0.46	-1.12	6.37	NO DIF	
Item 142	15.09	1.99	0.60	3.33	1.25	DIF	Favour rural examinees
Item 143	0.52	0.34	0.38	0.89	1.02	NO DIF	
Item 144	6.14	1.42	0.59	2.40	5.53	DIF	Favour rural examinees
Item 145	1.80	1.20	0.77	1.56	0.31	NO DIF	
Item 146	4.71	1.54	0.73	2.13	1.22	DIF	Favour rural examinees
Item 147	0.00	0.22	0.61	0.36	0.43	NO DIF	
Item 148	0.02	-0.22	0.83	-0.27	0.04	NO DIF	
Item 149	0.59	0.44	0.48	0.92	0.32	NO DIF	
Item 150	4.87	-1.11	0.46	-2.39	0.10	DIF	Favour urban examinees
Item 151	0.01	-0.03	0.39	-0.09	0.05	NO DIF	
Item 152	3.99	0.82	0.38	2.14	0.24	DIF	Favour rural examinees
Item 153	0.15	0.32	0.51	0.61	0.02	NO DIF	
Item 154	0.75	0.41	0.39	1.05	0.77	NO DIF	
Item 155	1.75	0.54	0.38	1.40	6.74	NO DIF	
Item 156	0.08	-0.31	0.55	-0.56	1.38	NO DIF	
Item 157	0.01	-0.10	0.53	-0.19	0.26	NO DIF	
Item 158	2.24	1.33	0.81	1.64	0.64	NO DIF	
Item 159	20.02	-1.36	0.33	-4.05	8.20	DIF	Favour urban examinees
Item 160	1.18	0.47	0.38	1.22	1.54	NO DIF	
Item 161	0.34	0.25	0.34	0.72	0.38	NO DIF	
Item 162	21.83	-2.28	0.57	-4.03	1.21	DIF	Favour urban examinees
Item 163	3.79	-2.20	1.16	-1.90	1.19	NO DIF	

Item 164	9.36	-1.12	0.37	-3.01	1.22	DIF	Favour urban examinees
Item 165	1.04	-0.65	0.53	-1.22	1.82	NO DIF	
Item 166	15.66	-1.39	0.36	-3.84	3.40	DIF	Favour rural examinees
Item 167	0.15	0.12	0.27	0.47	12.73	NO DIF	
Item 168	0.00	0.05	0.31	0.15	1.51	NO DIF	
Item 169	11.87	-1.23	0.36	-3.42	0.57	DIF	Favour urban examinees
Item 170	10.49	0.93	0.33	2.84	11.58	DIF	Favour rural examinees
Item 171	6.70	1.01	0.50	2.02	39.84	DIF	Favour rural examinees
Item 172	0.14	0.45	0.71	0.63	12.04	NO DIF	
Item 173	0.36	0.60	0.72	0.83	12.98	NO DIF	
Item 174	0.09	-0.24	0.47	-0.51	0.00	NO DIF	
Item 175	19.26	2.02	0.53	3.83	0.13	DIF	Favour rural examinees
Item 176	8.55	1.47	0.53	2.77	0.02	DIF	Favour rural examinees
Item 177	19.28	0.00	0.00	0.00	0.00	NO DIF	
Item 178	5.73	1.19	0.50	2.37	0.32	DIF	Favour rural examinees
Item 179	6.57	1.25	0.50	2.53	0.11	DIF	Favour rural examinees
Item 180	7.05	1.20	0.50	2.42	0.02	DIF	Favour rural examinees
Item 181	16.35	0.00	0.00	0.00	0.00	NO DIF	
Item 182	2.53	0.68	0.45	1.51	2.26	DIF	Favour rural examinees
Item 183	16.29	0.00	0.00	0.00	0.00	NO DIF	
Item 184	17.28	0.00	0.00	0.00	0.00	NO DIF	
Item 185	0.90	0.41	0.42	0.96	4.69	NO DIF	

Item 186	6.73	1.25	0.52	2.40	0.37	DIF	Favour rural examinees
Item 187	0.82	0.00	0.00	0.00	0.00	NO DIF	
Item 188	1.01	-0.95	0.76	-1.26	0.67	NO DIF	
Item 189	1.22	-0.73	0.58	-1.26	3.01	NO DIF	
Item 190	9.69	0.96	0.43	2.24	63.33	DIF	Favour rural examinees
Item 191	4.86	1.02	0.55	1.85	24.68	DIF	Favour rural examinees
Item 192	1.12	1.53	1.16	1.32	5.93	NO DIF	
Item 193	2.36	0.00	0.00	0.00	.	NO DIF	
Item 194	2.82	-1.15	0.68	-1.68	3.09	NO DIF	
Item 195	3.02	-2.09	1.14	-1.84	2.22	NO DIF	
Item 196	42.07	-2.85	0.74	-3.82	18.30	DIF	Favour urban examinees
Item 197	1.99	1.53	0.99	1.55	4.13	NO DIF	
Item 198	0.46	-0.36	0.41	-0.88	0.03	NO DIF	
Item 199	0.32	-0.30	0.39	-0.77	0.03	NO DIF	
Item 200	0.42	-0.33	0.39	-0.85	0.01	NO DIF	
Item 201	0.23	-0.27	0.39	-0.68	0.06	NO DIF	
Item 202	0.46	-0.34	0.39	-0.88	0.07	NO DIF	
Item 203	8.39	-1.07	0.36	-3.00	2.16	DIF	Favour urban examinees
Item 204	9.80	-1.15	0.36	-3.18	2.31	DIF	Favour urban examinees
Item 205	11.10	-1.19	0.36	-3.32	2.65	DIF	Favour urban examinees
Item 206	9.80	-1.15	0.36	-3.18	2.17	DIF	Favour urban examinees
Item 207	5.50	-1.13	0.52	-2.17	4.37	DIF	Favour urban examinees
Item 208	18.39	-2.01	0.60	-3.38	12.18	DIF	Favour urban examinees

Item 209	11.24	1.10	0.33	3.36	1.54	DIF	Favour rural examinees
Item 210	19.14	-1.28	0.33	-3.92	6.07	DIF	Favour urban examinees
Item 211	0.23	0.29	1.19	0.24	2.78	NO DIF	
Item 212	7.01	0.00	0.00	0.00	0.00	NO DIF	
Item 213	15.10	-1.28	0.35	-3.65	0.54	DIF	Favour urban examinees
Item 214	4.81	2.76	1.01	2.72	0.02	DIF	Favour rural examinees
Item 215	4.05	-0.93	0.46	-2.03	4.77	DIF	Favour urban examinees
Item 216	12.35	1.06	0.33	3.18	13.83	DIF	Favour rural examinees
Item 217	3.64	0.76	0.43	1.77	5.81	NO DIF	
Item 218	0.02	-0.07	0.51	-0.13	0.23	NO DIF	
Item 219	0.00	0.03	0.28	0.13	1.51	NO DIF	
Item 220	0.05	0.29	0.57	0.50	0.00	NO DIF	
Item 221	18.53	-1.57	0.39	-4.08	0.08	DIF	Favour urban examinees
Item 222	0.19	0.43	1.60	0.27	0.01	NO DIF	
Item 223	0.08	0.07	0.77	0.09	0.12	NO DIF	
Item 224	0.04	-0.57	0.93	-0.62	1.55	NO DIF	
Item 225	4.24	-0.77	0.38	-2.04	0.67	DIF	Favour urban examinees
Item 226	0.60	0.57	0.58	0.98	0.22	NO DIF	

Presents item parameters in the DEV-ERLT using IRT framework

ITEM	IRT	
	a	***b***
ERLT0001	0.936	-4.472
ERLT0002	1.177	-3.133
ERLT0003	1.574	-2.494
ERLT0004	2.175	-1.886
ERLT0005	2.832	-1.691
ERLT0006	2.553	-1.505
ERLT0007	3.225	-1.124
ERLT0008	3.183	-1.041
ERLT0009	4.058	-0.980
ERLT0010	2.894	-0.904
ERLT0011	4.94	-0.938
ERLT0012	4.678	-0.849
ERLT0013	4.732	-0.813
ERLT0014	5.477	-0.808
ERLT0015	4.944	-0.865
ERLT0016	4.647	-0.723
ERLT0017	4.502	-0.578
ERLT0018	4.984	-0.622
ERLT0019	4.421	-0.685
ERLT0020	4.838	-0.628
ERLT0021	6.352	-0.578
ERLT0022	0.668	-0.54
ERLT0023	2.02	-0.862
ERLT0024	3.065	-0.895
ERLT0025	1.478	-2.409
ERLT0026	3.912	-0.48
ERLT0027	3.928	-0.601
ERLT0028	1.109	-2.765
ERLT0029	2.14	-0.793

ERLT0030	5.242	-0.441
ERLT0031	3.379	-0.713
ERLT0032	3.429	-0.723
ERLT0033	3.981	-0.611
ERLT0034	4.009	-0.571
ERLT0035	3.938	-0.637
ERLT0036	5.409	-0.475
ERLT0037	0.636	-0.698
ERLT0038	4.525	-0.424
ERLT0039	1.812	-1.473
ERLT0040	1.629	-1.681
ERLT0041	3.118	-0.799
ERLT0042	2.911	-0.72
ERLT0043	0.627	-3.759
ERLT0044	0.914	-3.124
ERLT0045	1.035	-2.771
ERLT0046	1.642	-1.921
ERLT0047	2.001	-1.636
ERLT0048	1.533	-1.501
ERLT0049	2.344	-1.024
ERLT0050	2.593	-0.972
ERLT0051	2.433	-0.837
ERLT0052	2.987	-0.81
ERLT0053	4.035	-0.895
ERLT0054	2.872	-0.641
ERLT0055	4.036	-0.764
ERLT0056	4.372	-0.77
ERLT0057	4.641	-0.822
ERLT0058	4.193	-0.697
ERLT0059	0.785	-0.199
ERLT0060	2.24	-0.535
ERLT0061	2.677	-0.663

ERLT0062	3.728	-0.598
ERLT0063	4.459	-0.517
ERLT0064	5.148	-0.509
ERLT0065	3.533	-0.551
ERLT0066	0.71	-0.473
ERLT0067	3.06	-0.778
ERLT0068	2.153	-0.688
ERLT0069	1.385	-2.013
ERLT0070	3.486	-0.666
ERLT0071	3.334	-0.633
ERLT0072	2.616	-0.86
ERLT0073	5.02	-0.409
ERLT0074	3.961	-0.596
ERLT0075	3.398	-0.402
ERLT0076	4.513	-0.477
ERLT0077	1.236	-1.374
ERLT0078	2.753	-0.605
ERLT0079	3.662	-0.539
ERLT0080	3.801	-0.709
ERLT0081	3.683	-0.824
ERLT0082	1.526	-0.206
ERLT0083	3.01	-0.593
ERLT0084	2.605	-0.462
ERLT0085	2.354	-0.374
ERLT0086	0.52	-4.861
ERLT0088	0.191	-5.399
ERLT0089	1.56	-0.431
ERLT0090	2.412	0.399
ERLT0091	1.102	-2.27
ERLT0092	1.657	-0.398
ERLT0093	1.591	-0.268
ERLT0094	0.132	-6.901

ERLT0095	1.527	-0.795
ERLT0096	0.685	-1.465
ERLT0097	2.073	-0.528
ERLT0098	2.765	-0.648
ERLT0099	2.018	-0.855
ERLT0100	2.696	-0.587
ERLT0101	3.02	-0.672
ERLT0102	2.54	-0.592
ERLT0103	1.989	-0.224
ERLT0104	2.175	-0.606
ERLT0105	1.348	-1.414
ERLT0106	1.459	0.18
ERLT0107	1.581	-0.733
ERLT0108	1.599	-0.922
ERLT0109	2.208	-0.657
ERLT0110	2.027	-0.804
ERLT0111	2.456	-0.563
ERLT0112	2.463	-0.37
ERLT0113	2.169	-0.556
ERLT0114	1.668	-0.54
ERLT0115	2.93	-0.311
ERLT0116	1.811	0.146
ERLT0117	1.591	0.179
ERLT0118	0.942	-1.732
ERLT0119	1.055	-1.717
ERLT0120	1.682	-1.279
ERLT0121	1.207	-2.038
ERLT0122	1.494	-1.562
ERLT0123	2.179	-0.826
ERLT0124	2.062	0.227
ERLT0125	2.039	-0.777
ERLT0126	1.743	-0.955

ERLT0127	0.881	-0.481
ERLT0128	1.051	-0.327
ERLT0129	2.949	0.466
ERLT0130	3.782	0.454
ERLT0131	3.595	0.448
ERLT0132	1.977	-0.557
ERLT0133	2.125	-0.412
ERLT0134	1.708	-0.135
ERLT0135	1.669	0.043
ERLT0136	3.908	0.236
ERLT0137	3.241	-0.318
ERLT0138	3.514	0.212
ERLT0139	3.862	0.234
ERLT0140	4.12	-0.266
ERLT0141	3.283	0.037
ERLT0142	1.241	-0.224
ERLT0143	1.222	-0.216
ERLT0144	1.184	-0.197
ERLT0145	1.241	-0.187
ERLT0146	1.229	-0.229
ERLT0147	1.495	0.579
ERLT0148	0.464	0.949
ERLT0149	0.630	-0.835
ERLT0150	3.768	0.486
ERLT0151	4.480	0.489
ERLT0152	0.771	-0.930
ERLT0153	2.572	0.400
ERLT0154	0.824	0.649
ERLT0155	1.082	0.020
ERLT0156	1.971	0.407
ERLT0157	0.972	-0.853
ERLT0158	2.973	0.455

ERLT0159	3.277	0.527
ERLT0160	1.311	0.123
ERLT0161	1.164	0.764

Statistics of ability scores in the DEV-ERLT using IRT framework

Examinees	Ability score
1	-1.333
2	0.445
3	-1.333
4	-0.444
5	0.444
6	-1.333
7	0.444
8	0.444
9	-0.441
10	-0.410
11	-1.346
12	-0.473
13	-0.173
14	0.443
15	2.172
16	-0.444
17	0.444
18	-0.444
19	-0.444
20	0.445
21	0.428
22	-0.444
23	-0.444
24	-0.444
25	-0.444
26	-0.444
27	0.444
28	-0.444
29	-0.410
30	-0.444

31	0.444
32	-0.444
33	-0.444
34	-0.444
35	-0.444
36	-0.444
37	-0.444
38	-0.173
39	0.428
40	0.291
41	0.444
42	0.444
43	0.444
44	0.444
45	-0.444
46	-0.444
47	-0.444
48	-0.444
49	-0.444
50	-0.444
51	-0.444
52	-0.444
53	-0.444
54	-0.444
55	-0.444
56	-0.444
57	-0.444
58	0.517
59	-0.444
60	-0.444
61	-0.444
62	-1.333

63	-0.444
64	-0.444
65	0.444
66	0.428
67	-0.444
68	-1.333
69	-0.444
70	-0.444
71	-0.444
72	-0.444
73	-0.444
74	-0.444
75	-0.444
76	-0.410
77	0.444
78	0.444
79	0.291
80	0.444
81	0.444
82	0.444
83	0.444
84	0.445
85	0.452
86	0.444
87	-0.444
88	0.444
89	-0.444
90	-2.222
91	-2.177
92	-2.218
93	-2.177
94	-1.333

95	-1.333
96	-1.333
97	-3.019
98	-2.222
99	-2.222
100	-1.331
101	-1.335
102	-2.222
103	-1.896
104	-2.218
105	-1.333
106	-1.333
107	-1.346
108	-2.222
109	-2.223
110	-1.333
111	-2.218
112	-1.333
113	-1.333
114	0.444
115	0.444
116	0.444
117	0.444
118	0.444
119	0.444
120	0.444
121	0.444
122	0.444
123	0.444
124	0.444
125	0.444
126	0.444

127	0.444
128	0.444
129	0.444
130	0.443
131	0.444
132	0.444
133	0.444
134	0.444
135	0.444
136	0.444
137	-0.173
138	0.444
139	0.442
140	0.442
141	0.440
142	0.441
143	0.443
144	0.414
145	0.445
146	0.434
147	0.445
148	0.444
149	0.464
150	0.444
151	0.444
152	0.452
153	0.445
154	0.445
155	0.444
156	0.444
157	-0.441
158	0.444

159	0.444
160	0.444
161	0.444
162	0.444
163	0.444
164	0.444
165	0.444
166	0.444
167	0.444
168	0.444
169	0.291
170	0.444
171	0.445
172	0.444
173	0.444
174	0.444
175	0.444
176	0.291
177	0.444
178	0.444
179	0.444
180	0.291
181	0.444
182	0.444
183	0.428
184	0.444
185	0.444
186	0.444
187	0.444
188	0.444
189	0.444
190	0.444

191	0.444
192	0.444
193	0.444
194	0.444
195	0.444
196	0.444
197	0.444
198	0.444
199	0.444
200	-0.173
201	0.444
202	0.444
203	0.444
204	0.444
205	0.444
206	0.444
207	0.444
208	0.428
209	0.444
210	0.444
211	-0.444
212	0.444
213	-0.444
214	-2.222
215	-0.444
216	-0.444
217	-0.444
218	-1.333
219	-1.333
220	-1.333
221	-0.444
222	-1.333

223	-0.444
224	-0.444
225	0.444
226	-1.333
227	-1.333
228	0.444
229	-0.445
230	-0.444
231	-0.444
232	-0.444
233	-0.444
234	-1.333
235	-1.333
236	-1.333
237	-1.333
238	-0.444
239	-0.473
240	-1.335
241	0.291
242	-1.335
243	-0.444
244	-1.333
245	-0.444
246	-0.444
247	-1.333
248	0.444
249	0.444
250	-0.444
251	-1.455
252	-1.896
253	-1.455
254	-1.333

255	-1.333
256	0.291
257	-0.444
258	-0.444
259	-0.444
260	-1.335
261	-1.896
262	-0.444
263	-0.444
264	-0.473
265	-1.333
266	-1.333
267	-1.333
268	-1.333
269	-1.333
270	-1.455
271	-1.333
272	-2.222
273	-1.455
274	-0.683
275	-2.218
276	-0.444
277	-1.333
278	-1.333
279	-1.331
280	-1.333
281	-0.444
282	-0.444
283	-0.444
284	-1.333
285	-1.333
286	-1.333

287	-2.218
288	-1.155
289	-1.333
290	-1.333
291	0.444
292	0.444
293	-0.444
294	0.444
295	0.444
296	0.444
297	-0.173
298	0.444
299	0.444
300	0.444
301	0.443
302	0.444
303	0.444
304	0.291
305	0.444
306	0.444
307	-1.333
308	-1.333
309	-1.333
310	-1.333
311	-1.333
312	-1.333
313	-1.333
314	-1.333
315	-1.333
316	-1.333
317	-1.333
318	-1.333

319	-1.333
320	-1.333
321	-1.333
322	-1.333
323	-1.333
324	0.444
325	0.291
326	-0.444
327	0.444
328	0.444
329	-0.444
330	0.444
331	0.428
332	0.291
333	0.428
334	0.428
335	0.444
336	0.443
337	0.428
338	0.444
339	0.443
340	0.443
341	0.444
342	0.291
343	0.443
344	0.428
345	0.291
346	0.428
347	0.444
348	0.291
349	0.291
350	0.291

351	0.444
352	0.443
353	0.444
354	0.444
355	0.444
356	0.444
357	0.444
358	0.444
359	0.444
360	0.444
361	0.444
362	0.444
363	0.444
364	0.444
365	0.444
366	0.444
367	0.444
368	0.443
369	0.443
370	0.443
371	0.444
372	0.443
373	0.444
374	0.443
375	0.444
376	1.333
377	1.333
378	1.333
379	1.326
380	1.326
381	1.256
382	1.333

383	1.326
384	1.334
385	1.333
386	1.333
387	1.326
388	1.333
389	1.333
390	1.334
391	1.334
392	1.326
393	1.333
394	1.326
395	1.334
396	1.333
397	1.333
398	0.452
399	1.326
400	1.334
401	1.326
402	0.881
403	1.334
404	1.333
405	1.334
406	1.336
407	1.334
408	0.452
409	1.334
410	1.326
411	1.333
412	1.326
413	1.334
414	1.334

415	1.333
416	1.333
417	1.333
418	1.334
419	1.334
420	1.326
421	1.334
422	0.517
423	0.517
424	1.326
425	1.326
426	1.326
427	1.326
428	1.326
429	1.326
430	1.326
431	1.326
432	1.326
433	1.326
434	1.326
435	1.326
436	1.326
437	1.326
438	1.326
439	1.326
440	1.326
441	1.326
442	1.326
443	1.326
444	1.326
445	1.326
446	1.326

447	1.326
448	1.326
449	1.326
450	1.326
451	1.326
452	1.326
453	1.326
454	1.326
455	1.326
456	1.326
457	1.326
458	1.326
459	1.334
460	0.881
461	1.333
462	1.334
463	1.333
464	1.334
465	1.333
466	1.334
467	1.334
468	1.334
469	1.334
470	1.334
471	1.334
472	1.334
473	1.334
474	1.334
475	1.334
476	1.334
477	1.334
478	1.334

479	1.334
480	1.334
481	1.334
482	1.334
483	1.334
484	1.334
485	1.334
486	1.334
487	1.334
488	1.334
489	1.334
490	1.334
491	1.334
492	1.334
493	1.334
494	1.334
495	1.334
496	1.334
497	1.334
498	1.334
499	1.334
500	1.334
501	1.334
502	1.334
503	1.334
504	1.334
505	1.334
506	1.336
507	1.336
508	1.336
509	1.336
510	1.336

511	1.336
512	1.336
513	1.336
514	1.326
515	1.336
516	1.334
517	1.334
518	1.334
519	1.334
520	1.334
521	1.334
522	1.334
523	1.334
524	1.334
525	1.334
526	1.334
527	1.334
528	1.334
529	1.336
530	1.334
531	1.334
532	1.336
533	1.336
534	1.334
535	1.334
536	1.334
537	1.336
538	1.336
539	1.334
540	1.334
541	1.334
542	1.334

543	1.334
544	1.334
545	1.334
546	1.334
547	1.334
548	1.334
549	1.334
550	1.336
551	1.336
552	1.336
553	1.334
554	1.334
555	1.334
556	1.334
557	1.336
558	1.334
559	1.336
560	1.334
561	1.336
562	1.334
563	1.334
564	1.334
565	1.334
566	1.334
567	1.336
568	1.336
569	1.334
570	1.334
571	1.334
572	1.334
573	1.334
574	1.334

575	1.334
576	1.334
577	1.334
578	1.336
579	1.336
580	1.334
581	1.334
582	1.334
583	1.334
584	1.334
585	1.334
586	1.334
587	1.334
588	1.334
589	1.334
590	1.336
591	1.336
592	1.334
593	1.334
594	1.336
595	1.336
596	1.336
597	1.326
598	1.336
599	-1.333
600	0.881
601	0.881
602	0.881
603	0.881
604	0.881
605	0.881
606	1.256

607	1.256
608	1.256
609	1.256
610	1.256
611	1.256
612	1.256
613	1.336
614	1.336
615	1.336
616	1.334
617	1.366
618	-1.333
619	1.336
620	1.336
621	1.336
622	1.336
623	1.336
624	1.336
625	1.336
626	1.336
627	1.336
628	1.336
629	1.326
630	1.336
631	1.336
632	-1.333
633	-0.444
634	-0.683
635	-0.444
636	-1.333
637	-1.333
638	-1.333

639	-0.444
640	-1.333
641	-1.333
642	-1.333
643	-0.444
644	-0.444
645	-1.333
646	-1.333
647	-1.333
648	-1.333
649	-0.444
650	-1.333
651	-1.333
652	-1.333
653	-1.333
654	-1.155
655	-1.333
656	-1.333
657	-0.444
658	-0.444
659	-1.333
660	-1.155
661	-1.333
662	-1.333
663	-0.683
664	-1.346
665	-2.177
666	-1.333
667	-1.333
668	-1.333
669	-0.444
670	-0.683

671	-1.333
672	-0.444
673	-1.333
674	-0.444
675	-1.333
676	-0.473
677	-1.333
678	-1.333
679	-1.333
680	-0.444
681	-1.333
682	-1.333
683	-0.444
684	-1.333
685	-1.333
686	-1.333
687	-0.444
688	-1.333
689	-0.444
690	-1.333
691	-0.444
692	-1.333
693	-1.333
694	-1.333
695	-0.444
696	-0.444
697	-0.444
698	-1.333
699	-0.444
700	-1.333
701	-1.333
702	-1.896

703	-1.335
704	-1.896
705	-2.222
706	-1.335
707	-1.333
708	-1.333
709	-1.333
710	-1.333
711	-1.333
712	-1.455
713	-1.333
714	-1.333
715	-1.333
716	-1.346
717	-2.218
718	-2.222
719	-2.223
720	-2.222
721	-2.222
722	-2.222
723	-2.222
724	-2.222
725	-2.222
726	-1.333
727	-1.333
728	-1.333
729	-1.333
730	-1.333
731	-0.444
732	-1.333
733	0.444
734	-0.444

735	-1.333
736	-1.333
737	0.291
738	-0.444
739	-1.333
740	0.444
741	-0.444
742	-1.333
743	-1.333
744	-0.444
745	-2.222
746	-1.455
747	-0.444
748	0.428
749	-0.444
750	-0.441
751	-1.333
752	-0.444
753	-0.444
754	-0.444
755	-0.444
756	-0.444
757	-0.444
758	-1.333
759	-1.333
760	-1.333
761	0.444
762	-1.333
763	-0.444
764	0.444
765	-1.335
766	-2.222

767	-1.333
768	-1.333
769	0.423
770	-1.203
771	0.452
772	-1.352
773	-1.335
774	-0.444
775	-2.218
776	-0.444

Difficulty Levels of Identification of Objects, Animals, Fruits and Parts of Human Body

ITEM	IRT	
	A	B
IDEN01	1.526	-0.206
IDEN02	3.01	-0.593
IDEN03	2.605	-0.462
IDEN04	2.354	-0.374
IDEN05	0.52	-4.861
IDEN06	0.191	-5.399
IDEN08	1.56	-0.431
IDEN09	2.412	0.399
IDEN10	1.102	-2.27
IDEN11	1.657	-0.398
IDEN12	1.591	-0.268
IDEN13	0.132	-6.901
IDEN14	1.527	-0.795
IDEN15	0.685	-1.465
IDEN16	2.073	-0.528
IDEN17	2.765	-0.648
IDEN18	2.018	-0.855
IDEN19	2.696	-0.587
IDEN20	3.02	-0.672

IDEN21	2.54	-0.592
IDEN22	1.989	-0.224
IDEN23	2.175	-0.606
IDEN24	1.348	-1.414
IDEN25	1.459	0.18
IDEN26	1.581	-0.733
IDEN27	1.599	-0.922
IDEN28	2.208	-0.657
IDEN29	2.027	-0.804
IDEN30	2.456	-0.563
IDEN31	2.463	-0.37
IDEN32	2.169	-0.556
IDEN33	1.668	-0.54
IDEN34	2.93	-0.311
IDEN35	1.811	0.146
IDEN36	1.591	0.179
IDEN37	0.942	-1.732
IDEN38	1.055	-1.717
IDEN39	1.682	-1.279
IDEN40	1.207	-2.038
IDEN41	1.494	-1.562
IDEN42	2.179	-0.826
IDEN43	2.062	0.227
IDEN44	2.039	-0.777
IDEN45	1.743	-0.955