

https://dspace.org/

A pilot validation study of the OLSET Picture-Language Vocabulary Test.

Malope, Nomxolisi Sindiswa

2011-04-21
http://hdl.handle.net/10539/9616

Downloaded from a DSpace repository. https://dspace.org/

A pilot validation study of the OLSET Picture-Language Vocabulary Test

Nomxolisi Sindiswa Malope

Department of Psychology

University of the Witwatersrand,

Private Bag x1

Wits

2050

A dissertation submitted to the Faculty of Humanities, University of the Witwatersrand, Johannesburg, in partial fulfilment of the requirements of the degree of Masters of Arts in Psychology by Research and Coursework.

Johannesburg, 2009

DECLARATION

I declare that this dissertation is my own unaided work. It is submitted for the degree of Masters in Psychology by Research and Coursework in the University of the Witwatersrand, Johannesburg. It has not been submitted before under any other degree or examination in any other university.

Signed: _____on:_____day of_____2010 at _____.

ABSTRACT

This study was aimed at establishing the psychometric properties of the OLSET Picture-Language Vocabulary Level 1 Test (OPLVLT) which is a group test developed in South Africa for children in the lower primary school. The test has been used in a number of studies conducted in second language classrooms, but little work has been conducted on its standardisation.

The standardisation sample was a sample of convenience, comprising a hundred and twenty (120) Grade One Learners drawn from three schools in Johannesburg. Most children were second language speakers of English, and were aged between 6 and 11 years of age.

Reliability of the OPLVLT was established by calculating Cronbach alpha which yielded a value of 0.86. Each of the 36 Items had alpha's ranging between 0.83 and 0.85 when deleted and thus they were all retained for the test. The construct validity of the OPLVLT was then investigated by correlating scores on the test with scores on an individual test (the Peabody Picture Vocabulary Test PPVT-R) and another group test (the Metropolitan School Language and Listening subtest MRT). These instruments have been widely used internationally, and have well-established reliability and construct validity.

The correlation between the OPLVLT and PPVT was 0.17. The correlation between the OPLVLT and MRT was 0.29. Given the low relationship between the tests, an overall reliability estimate for the construct validity analysis was calculated. The Cronbach's alpha for this analysis was 0.57, indicating a low level of internal consistency between the different tests chosen for use in the construct validation.

Given its high internal reliability as a test, but its low level of correlation with other tests of language ability, the overall conclusion from this pilot study was that the OLSET Picture-Language Vocabulary Level 1 Test was a reliable test which was not highly correlated with either the Peabody Picture Vocabulary subtest or the Metropolitan School Language and Listening subtest in this particular sample. The high internal reliability indices would suggest that the OPLVLT has potential as a psychometric instrument.

ACKNOWLEDGEMENTS

I am indebted to the following people in the successful completion of this project and would like to express my heart-felt gratitude:

- Professor Charles Potter my supervisor for your advice and assistance in this project.
- Peter Fridjhon for your statistical expertise on the design and analysis of the project.
- Sabrina Liccardo for your continued support throughout the project.
- The staff at the Open Learning Education Systems Trust for availing yourselves and availing their data. I could not have done this without you.
- All the Grade 1 learners of 2008 at Welizibuko, Paul Mosaka and Bree Primary Schools, you are South Africa's bright future!
- The Third year Psychology students of 2008, who assisted with the data collection.
- Finally, My family and friends; for your continued faith in me, even when I had none in myself.

TABLE OF CONTENTS

DECLARATION 2

ACKNOWLEDGEMENTS 4

TABLE OF CONTENTS 5

LIST OF FIGURES 9

LIST OF TABLES 9

LIST OF GRAPHS 10

ABBREVIATIONS 11

CHAPTER 1 12

INTRODUCTION 12

1.1. Psychometric Testing and Assessment in a Multicultural Society 12

1.2. Focuses of this study 12

1.3. Assumptions 12

1.4. The South African Education System 13

1.5. The Use of Testing for Monitoring, Evaluation and Quality Improvement 14

1.6. How is South African Education Currently Rated Internationally? 14

1.7. Aims of this Study 15

1.8. Rationale 16

1.9. Research Questions 16

CHAPTER 2 18

GENERAL ORIENTATION, PROBLEM FORMULATION AND RELEVANCE OF THIS STUDY 18

2.1. Problem Investigated 18

2.2. Variables in the Study 18

2.3. Relevance of these Variables to South African Education 18

2.4. Language Teaching Policy in South Africa 19

2.5. Teachers 20

2.6. Learners....21
2.7. Departments of Education....21
2.8. The Open Learning Systems Education Trust (OLSET)....22
CHAPTER 3....24
LITERATURE REVIEW....24
Introduction....24
PART A....24
CONSTRUCT VALIDATION....24
3.1. Defining Test Validation....24
3.2. The Process of Construct Validation....26
PART B....28
VOCABULARY....28
3.3. Vocabulary Defined....28
3.4. Knowing a word....29
3.5. Vocabulary Acquisition....30
3.6. The importance of vocabulary....31
3.7. Vocabulary Instruction....32
3.8. Vocabulary Instruction for young children....33
3.9. Vocabulary Assessment....37
CHAPTER 4....39
RESEARCH DESIGN AND METHODOLOGY....39
4.1. Research Questions Guiding the Study....39
4.2. Research Design....41
4.3. Instruments used in this Study....42
4.4. Sample....44
4.5. Data Collection Procedure....47
4.6. Ethical Considerations....48

4.7. Summary Statistics....48
4.8. Reliability Analyses....50
4.9. Validity Analyses....51
4.10. Group Differences Analysis....52
CHAPTER 5....56
FINDINGS....56
5.1. Reliability....57
5.1.1. Internal Consistency....58
5.1.2. Item Analysis....59
5.1.3. Item Discrimination....62
5.1.4. Item Difficulty....63
5.2. Validity....64
5.2.1. Construct Validity....64
5.3. ANOVA....66
5.3.2. Non-Parametric ANOVA....67
5.5. Summary of results....73
CHAPTER 6....75
DISCUSSION....75
6.1. Extraneous variables that may have influence the test scores;....75
6.1.1. The formats of the tests....75
6.1.2. Socio-Economic Status....77
6.1.3. Test Anxiety....78
6.1.4. Studies of group differences among test-takers....79
6.2. Implications of the test as it stands....80
6.2.1. The construct measured by the test....80
6.2.3. Discrimination power of the OPLVLT....80
6.3. Limitations....81

6.4. Recommendations 83
6.4.1. Recommendation for Vocabulary teaching 83
6.4.2. Factor of Language 83
6.4.2. Recommendations for Test Developers 83
CHAPTER 7 85
CONCLUSION 85
REFERENCES 88
APPENDICES 101
Appendix A 101
Pilot Standardisation Sample 101
Appendix B 102
Background on the Organisation 102
Procedures used by OLSET to get subjects and administer tests 102
Appendix C 103
Metropolitan Readiness Level 1 Test: School Language and Listening Subtest 103
Concurrent Validity coefficients 103
Appendix D 104
Peabody Picture Vocabulary Test 104
Split-half Reliability coeffiecient for Form L. 104
Appendix E 105
OLSET Picture-Language Vocabulary Level 1-Test 105
Appendix F 106
Item Discrimination Calculations 106
Appendix G 107
Scores attained by learners speaking different languages 107
Appendix H: 108
Test for Normality of OPLVLT Data: Probability plots 108

LIST OF FIGURES

Figure 4.1: Frequency of OPLVLT Scores..... 50

Figure 4.2: Cronbach's alpha illustration.....51

Figure 5.1: Statistical Decision Tree.....69

LIST OF TABLES

Table 4.1: Summary of School Frequencies.....45

Table 4.2: Gender Frequencies.....45

Table 4.3: Distribution of learners' home languages.....47

Table 4.4: Means of scores attained on tests.....49

Table 5.1: Research questions, statistical techniques and variable measured.....57

Table 5.2: Cronbach's alpha for Internal reliability.....59

Table 5.3: Inter-Item correlations and corresponding alpha's.....61

Table 5.4: Item Difficulty.....63

Table 5.5: Summary of scores on the validation tests.....65

Table 5.6: Cronbach Coefficient alpha.....66

Table 5.7: Descriptive Statistics for Gender as an IV…….....67

Table 5.8: Non-parametric ANOVA Results……...67

Table 5.9: Non-parametric significance for Gender…….....68

Table 5.10: Non-parametric Kruskal-Wallis for Gender.....70

Table 5.11: ANOVA for Gender and Age……....71

Table 5.12: Strength of the effects on score…….....72

Table 5.13: ANCOVA for Interaction of age and gender…….....72

Table 5.14: Non-parametric ANOVA for Age and Gender as separate variable if the data were normally distributed.....72

Table 5.15: Parameters for Age and Gender.....73

Table 6.1: Normed scores on the sample.....81

Table 6.2: Scores on tests by Language.....82

LIST OF GRAPHS

Graph 4.1......46

ABBREVIATIONS

AERA: American Educational Research Association

ANCOVA: Analysis of Covariance

ANOVA: Analysis of Variance

APA: American Psychological Association

DfID: Department for International Development

HPCSA: Health Professions Council of South Africa

MSRT (SLL): Metropolitan School Readiness Subtest: School Language and Listening Test

NEPI: National Effective Parenting Initiative

OLSET: Open Learning Systems Education Trust

OPLVLT: OLSET Picture-Language Vocabulary Level 1 Test

PPVT-R: Peabody Picture Vocabulary Test (R=Revised; III=Third Edition)

SAIRR: South African Institute of Race Relations

SANPAD: South Africa Netherlands Research Programme on Alternatives in Development

SAQA: South African Qualifications Authority

USAID: United States Agency for International Development

WISC-R: Weschler Intelligence Scale for Children-Revised Edition

CHAPTER 1

INTRODUCTION

1.1. Psychometric Testing and Assessment in a Multicultural Society

The majority of Psychometric tests in general usage in South Africa have been normed on English speaking population groups (usually from the USA or Europe). As South Africa is a multicultural country which is highly diverse in terms of ethnic and socio-economic differences between communities, it is difficult to establish a basis for fair appraisal (Pereira, 2008). To do so would require the use of validated psychometric tests, as well as the use of locally developed tests. There are also issues of culture, for the reason that psychometric tests measure acquired knowledge, which is gained through participation in a dominant culture (Utley, Haywood and Masters, 1992).

1.2. Focuses of this study

This study focused on a validation of a group test of vocabulary which was developed in South African primary schools. Despite having been used for a number of years as a measure of English language proficiency, there had not, at the time of this research, a formal standardisation study on the instrument.

This research was a pilot study which examined the psychometric qualities of this instrument when used with a sample of Grade One learners who speak English as a second language. The study was a pilot study as the standardisation sample was a specific one which cannot be regarded as representative of the conditions pertaining in all South African communities and all South African primary schools. The aim of the research was to establish whether the test has adequate reliability, and whether it is highly correlated with other tests measuring developmental language abilities in children. If this was found to be the case, standardisation on broader and more representative samples would then be indicated and a testing manual developed.

1.3. Assumptions

The process of validating a test normally involves using a new instrument in conjunction with a number of other well-established psychometric tests with a representative sample to establish the reliability and validity of the instrument (Cronbach, 1988). In this study, the desired outcome was to validate a South African developed test which has been in use for a

number of years. This would be done by establishing its psychometric properties, so as to improve the test by amending the test items.

The focus of this study thus lay on providing evidence of the validity and reliability of the test and describe the properties of the distribution of scores in vocabulary obtained for the particular standardisation sample tested. Additional possible third variables influencing the data (gender, age) were also included in the analysis.

The reason for focusing on these aspects is that when a new test has been validated in this way, it is then assumed that it is possible and statistically appropriate to administer it to wider samples. In addition, after a new test has been validated and standardised it would allow for the reporting of an individual testee's position relative to a particular normative sample. This is done by reporting on sampling factors influencing the data, prior to attempting to estimate grade equivalents or normalised standard scores.

These are the basic assumptions on which this study was based. They form the basis for the methodology used in the study, which will be described in more detail later in this report, in Chapter Four.

1.4. The South African Education System

This study was conducted with learners at Grade One level. As tests of English language proficiency are needed to establish learning needs as well as learning gains in South African primary schools, the results of this study have relevance to South African education.

In the latest available South African records (South African Institute of Race Relations, (SAIRR), 2008) there were 1 172 659 learners enrolled in Grade One across the country. Forty-four percent of these learners had not had Early Childhood Education (pre-school or kindergarten).

In Gauteng province, where this study was conducted, there were 1 988 schools, constituting 8% of all schools in South Africa. There were 52 451 teachers and 1 710 851 learners, making the pupil-teacher ratio 33:1. This ratio is reasonably good as the standards stipulated by the South African government recommend a limit of 35 learners per teacher in a classroom.

However, South African school education is considered of a poor quality, as both internal and external evaluations attest (South African Institute of Race Relations, (SAIRR), 2008). This

largely derives from weakly performing schools and this raises severe concerns about the learners' further education, employment, earnings, careers and life chances generally as many studies have indicated that vocabulary is important for general success in life.

Absence of quality education therefore, is inarguably detrimental to the future of South African learners. The school system in South African is ranked number 50 in the World, 57 in the countries reaching the achievements of universal education, 39 on the Quality Input which indicates whether teaching and learning is given the necessary support, adequate infrastructure, teaching material and teacher training (SAIRR, 2008).

1.5. The Use of Testing for Monitoring, Evaluation and Quality Improvement

This background has been provided as this study is relevant to South African education, as well as to the local and international literature on use of testing for monitoring, evaluation and quality improvement in schools. Internationally, standardised tests are being increasingly used for purposes of gauging quality of education and of schooling. While this practice is not yet wide-spread in South African schooling, this may well become a feature of South African education in the future, particularly if standards in education do not improve despite the large amounts of public funding which are currently being diverted into the education system (Volante, 2005).

The reason for suggesting this is that South African education is increasingly being assessed on international standards, as well as on standards pertaining to more developed countries such as the United States of America or England. This has a number of implications, the most important being that there are differences between the norms of those countries to those of South Africa. There is therefore a need in South Africa, as in other countries, to have tests based on norms derived from learners in South African schools.

1.6. How is South African Education Currently Rated Internationally?

Current international statistics (SAIRR, 2008) indicate that South Africa is not ranked highly among the world's best countries with regards to education. There is thus potential and room for growth and improvement. To demonstrate this will require appropriate instruments, which are valid to make these types of comparisons.

This study assumes that locally standardised tests will be necessary to establish whether growth and improvement are taking place. This study is thus one of a number which will be necessary to provide the types of locally standardised instruments which will be necessary if

one plans to make any justified interpretations or come to conclusion on the basis of psychometric test information. Having a norm on which to gauge a test score inevitably requires some set of standards on which to base interpretation. It also requires test results produced from tests which are valid, and capable of reflecting the abilities of South African children of school-going age.

1.7. Aims of this Study

The central aim of this study was to establish the psychometric qualities of a new group test which was designed to be used in South African classrooms to measure the level of vocabulary of Grade One learners. If it could be shown to have adequate reliability and construct validity, the test scores yielded by this test could then be used to establish the level of developmental language skills of learners in the classroom. This information could then be used to inform decisions concerning instructional interventions, the necessity of these interventions, as well as classroom progress made by learners.

This study has three specific aims;

The first aim is to validate the OLSET Picture-Language Vocabulary Level 1 Test by providing evidence of its reliability and validity properties and conducting Item Analysis whose results would inform the retaining or discarding of some of the items.

The second aim was to estimate the influence of possible third variables on the data, relative to the possibility of using the data set to establish standardised scores for the sample population. The reason for doing so was that if third variables such as gender, language and age have not affected the data, the distribution of scores from the standardisation sample could then be used to provide some indication of the range of vocabulary knowledge among other learners with similar characteristics to the sample population (Aiken & Groth-Marnat, 2006).

As the standardisation sample used in this research was a small one, it was likely that the conclusions from this study include making recommendations for a bigger, national study which will be aimed at standardising the OPLVLT with a bigger, more representative sample, the third aim of the study. These recommendations would be informed by the results of this pilot study, and whether there is evidence that the new group test of vocabulary has the psychometric qualities necessary for its use as a basis for identifying best practices in teaching vocabulary in this particular sample, as well as more widely in schools.

1.8. Rationale

This study is important not only to South African education and schooling, but also for the more general reason that it is imperative to systematically investigate psychometric properties of all scales used in South African settings (HPCSA, 2002). The *Standards for Educational & Psychological Testing* (APA, 1985) recommends that evidence of validity need to be presented for the types of inferences for which the use of a test is recommended. It is only after the psychometric properties have been investigated that a test can be administered and results generated by it can inform any decisions or interventions.

As the OPLVLT is a new group test which has not previously been validated, this study thus had the potential of filling current knowledge gap. If the instrument would be found to be both reliable and valid, this research would contribute to the South African education by making available a locally standardised vocabulary scale. This can then be used to assist interventions in the classroom, both for informing instructional decisions as well as for progress monitoring.

1.9. Research Questions

This study was guided by five research questions, as follows:

Research Question 1: Is the OPLV Level 1 Test a reliable measuring instrument which produces consistent and accurate scores?

Research Question 2: Does the OPLV Level 1 Test measure vocabulary? Is the Test a valid measure of the construct; vocabulary, as it claims to be?

Research Question 3: Are there significant differences between the scores of male and female learners on the test?

Research Question 4: Are there significant differences between the scores of learners who speak English as a first language and those who speak other languages as first language on the test?

Research Question 5: Are there significant differences between the scores of learners varying with their age?

There were also a number of specific hypotheses which related to the aims of the study and specific aspects of the analysis, which are outlined in Chapter Four. Other third variables

such as differences in the test formats which possibly influenced the scores from the tests used in this study are also included in the hypotheses outlined in Chapter Four.

In summary, this research was psychometric in nature and was a validation study of the OLSET Picture-Language Vocabulary Level 1 Test (OPLVLT). The OPLVLT is a group test which can be administered to a group of learners in one sitting. It has been developed and amended over a number of years (three times between 1990 and 2007).

The instrument has been designed to measure the vocabulary knowledge of the English Vocabulary among Grade 1 learners in South Africa (hence 'Level 1') both first language English speakers and those who speak other languages as a first language besides English. The reason for this is that vocabulary developmental is a critical aspect of second language proficiency development and needs to be supported and monitored (Nation, 2006).

The chapters in the rest of this report are structured as follows:

Chapter Two provides the problem formulation, as well as a summary of each of the other psychometric tests which will be used in this study as a basis for construct validation of the OPLVL as an instrument. Relevant literature such as issues involved in test validation has been consulted, and an overview is provided in Chapter Three.

The procedures followed in validating the test and the rationale behind each of them are outlined in Chapter Four. Chapter Five presents the results relating to the reliability and validity analyses, as well as a summary of the item analyses. Chapter Six discusses the results in relation to the hypotheses, and links these back to psychological theories as well as factors that may have influenced the scores of the particular standardisation sample on the test.

The final chapter, Chapter Seven, then provides comment on the overall results, and draws conclusions as to whether the OPLVLT is a reliable and valid test which can be more widely used in South Africa as a measure of vocabulary. Limitations of this study are outlined, and suggestions made for further research and validation of this instrument on a wider scale.

CHAPTER 2

GENERAL ORIENTATION, PROBLEM FORMULATION AND RELEVANCE OF THIS STUDY

2.1. Problem Investigated

The problem investigated in this research was whether it is possible to validate the OLSET Picture Language Vocabulary Test through examination of its psychometric properties. The wider problem is that there is a dearth of knowledge about vocabulary testing in South Africa, and a current dearth in availability of standardised group vocabulary assessment scales in South Africa (Blachowicz, Fisher & Ogle, 2006).

2.2. Variables in the Study

The design for this study was non-experimental in that there was no manipulation of any of the variables in the study. There are three variables, and the aim was to establish the relationship between these. The first variable is a set of scores on the OLSET Picture Language Vocabulary Level 1 Test (OPLVLT). The other two variables are scores on two research instruments used alongside the OPLVLT. The first is a subtest of the Metropolitan School Readiness Test (the School Language and Listening Test. The second is the Peabody Picture-Vocabulary Test (Third Edition).

Other variables investigated in this study were gender, age and the home language spoken by the learners in the sample, and whether there are differences in performance on the tests used in this study which can be attributed to these variables.

2.3. Relevance of these Variables to South African Education

The research is relevant to South African education for the reason that South Africa is a culturally diverse country with eleven recognised languages. Given the emphasis on use of English as a language of learning and teaching from the early stages in primary school, the development of a reliable and valid group test of English language vocabulary is a high priority.

It is also necessary in terms of psychometric theory to systematically investigate the psychometric properties of all scales used in South African settings (HPCSA, 2002). There is a current shortage of validated scales for measuring language as well as cognitive constructs

relevant to schooling (Knuver & Brandsma, 1993). Without reliable and valid measures the consequence is an inability to accurately estimate the response of learners to instruction, and consequently, the failure to design and implement appropriate strategies to address identified gaps in the educational system.

An analogy would lie in medicine, where the inability to accurately identify the presence of the Human Immuno Virus in a person, renders any medical provider unable to diagnose and prescribe any intervention be it medicinal or a lifestyle change, thus, before an accurate identification, diagnosis, and intervention there precedes an instrument that has been shown to be a reliable and valid estimate of that construct which is utilised for the function of identifying any problem. This study aims to validate the instrument which is claimed to be measuring *vocabulary*. As will be further discussed in the Literature, Vocabulary can be defined as either the set of words that are understood by a person or the set of words likely to be used by that person when constructing new sentences (http://en.wikipedia.org). The word web defines it simply as a language user's knowledge of words in that language (Word Web, 2007). In this study, the vocabulary specifically focused on is English amongst Grade 1 learners.

There is a growing recognition in the field of language teaching, that vocabulary development is a vital aspect of children's language development (Finocchiaro. 1969). Vocabulary development has been shown to be closely linked to children's progress in school and in their accomplishment in learning to read. Generally, the ability to use English is widely recognised as a prerequisite to success in work, business, and higher education and also as important for successful writing and learning at university (Dunn & Dunn, 1981). It is essential, therefore, that all children have regular opportunities to learn new words and to incorporate these words into their daily conversations (Wasik, 2006).

2.4. Language Teaching Policy in South Africa

In South Africa, the high prestige of English, the negative social meaning of the African languages in high-function public contexts and the impracticality of using many different languages has led to a strong preference for English as language of learning and teaching. This preference is evident with the complete use of English as a language of learning and teaching in 83.33% of South African Universities and the remaining 16.67 use both English and Afrikaans) (SANPAD, 2008).

The policy adopted in South African universities in which English is the official language of learning and teaching presents a serious problem since black learners' English-language proficiency in South Africa is often not adequate for using it as language of learning, in such cases English acts as an obstacle to educational development (Dunn & Dunn, 1981). There is thus a need for the language of learning and teaching to be a language which learners know very well and this requires early planning and relevant programmes that will help the development of this language as early as the first grade of school. Prior to this planning and intervention a tool/instrument of measuring this is necessary, hence this study.

Factors such as language, socio-cultural affiliation, and economic status affect psychometric test performances. Psychometric test scores are considered important indicators of competence in many institutions and works of place. Thus, if a psychometric test-taker does not understand the language in which the test is, it becomes almost impossible for the test-taker to do well on that test. However, although this study focuses on just language, specifically vocabulary, it is important to acknowledge that there are many other issues regarding tests such as the testee's acculturation, if the acculturation of the test-taker differs from those represented in the normative sample this may lead to biased and less valid results.

The results of this study have implications for a number of people and institutions. Below is a list of implications that this study might have.

2.5. Teachers

Teachers are most instrumental in the learning of children. They are the tools through which children learn. Their innovativeness and creativity in the classroom are most evident in the progress of their learners. Teachers receive a clear indicator of how much the children know, through the use of validated tests and they can utilise this information to design supplementary teachers resource materials which will promote vocabulary learning customised specifically for the needs of the learners in their classrooms (Savignon, 2006). Designing programmes to enhance teaching depend on a negotiation between researchers, policy makers and teachers, thus after teachers receive feedback from the researcher, they can in turn bring up suggestions in their Board meetings, and raise this with policy-makers and eventually be equipped with designing programmes that will benefit the learners in that way teachers are empowered to respond to local issues (that is, classroom challenges) appropriately.

2.6. Learners

Savignon (2006) describes how identifying the needs (both linguistic and social) using the method called Communicative Language Teaching assisted with designing the appropriate interventions for the immigrants and guest workers in Europe. In the same light, using validated testing instruments which are culture-free and normed on a similar sample allows tests to accurately reflect the challenges and thus assist learners through the development of a syllabus which is based on their needs. Learners also have a renewed interest in their learning when they discover what they know and what they are expected to know (Savignon, 2006). Hopefully, learners will also be encouraged to learn the words of the pictures that they did not recognise, and this study would have then been instrumental in further encouraging learning.

2.7. Departments of Education

Both local and national Departments of Education of South Africa have undergone a number of changes since the first democratic elections in the country in 1994. The education system has also undergone dramatic changes and so have the managerial tasks of education managers have also changed significantly resulting in a change in learning, growth, renewal and organisational development (Linde, 2006). These changes pose either a complex, 'great deal of upset' or success for the education system of South Africa (Fullan, 1993: vii).

Some of these changes were documented in the NEPI Report of 1992, The ANC Policy Framework of 1994, the Education White Paper of 1995 and 1996, the SAQA Act of 1995, The South African constitution, and the National Qualifications Framework. These documents outlined some critical changes such as (i) declaring that education is a human right, (ii) the state has a crucial role to play in the governance of school, (iii) that the state has an obligation to advice parents about rendering appropriate care and educational services to young children, (iv) open access to education and training opportunities of good quality for all children, (v) emphasise the redress of past inequalities, and (vi) they will provide quality education.

Both regional and national Departments of Education are likely to benefit from an instrument that allows teachers in classrooms to measure, accurately, the level of vocabulary that each child possesses and what they are able to do with this language. Improvement strategies that will assist with attaining the changes outlined above will be possible when there are

interventions informed by assessments of learner needs and specified end results (in this case, age and grade appropriate mastery of the English language).

Thus, this research is relevant for the reason that South Africa is a culturally diverse country which places a high value on use of English as a language of learning and teaching from the early stages in primary school. The development of a reliable and valid group test of English language vocabulary is thus a high priority.

The test investigated in this research has been used in the foundation phase of South African primary schools for a number of years. If it can be demonstrated to have high reliability and to be construct valid, the test could be useful to many teachers and schools, as well as to administrators and educational officials.

2.8. The Open Learning Systems Education Trust (OLSET)

OLSET was established in 1992, initially funded by the USAID and technical support was provided by LearnTech (Potter and Naidoo, 2006). Prior to the 1990's English was only introduced in Grade Three in South Africa's non-white schools, teaching in Grade One and Two was conducted in mother tongues (Rose & Tunmer, 1975). Since the 1990's to date English is regarded by many South Africans as a major international language. As a result, "most South African parents and learners feel that it is crucial for the learners to develop the ability to read, write, listen to and speak English in order to develop their potential" (Madileng, 2007: 1). In fact, consensus regarding the need for systematic and extensive vocabulary development is unanimous among authorities in reading instruction (Pany, Jenkins & Schreck, 1982: 202) and although there are Eleven (11) official languages in the country, the language of instruction in higher learning institutions is English and even in higher primary and high school instructions learners need to have a competent understanding of English. OLSET's radio interactive learning programme 'English in Action' thus set out to assist in the development of vocabulary from a very young age.

'English in Action' aims to provide a rich environment with active participation for learning English in a fun, creative, and critical thinking manner. They also aim to build up on learners' own practices and experiences and promote language development for personal expression, expose children to multiple models of natural English and foster effective listening, speaking, reading and writing skills (Naidoo & Potter, 2005). The programme uses carefully developed radio or audio cassette lessons to present natural language in dramatized situations to promote

specific learning activities and outcomes in the line with the government's goals. The initial model was focused on 'enhancing learner involvement and learner gains to a model of distance education and open learning focused on promoting teacher and learner gains through school, classroom and teacher support, and through in-service teacher training' with the aim of improving the quality of teaching in primary schools through developing English language competencies in junior primary phase (Naidoo & Potter, 2005). The methodology used emphasised 'fluency' as opposed to 'accuracy', 'language functions' as opposed to 'grammatical forms', 'balance between language and input, practice and communication output' as well as 'an emphasis on teaching students to become communicatively effective and to be able to use English in the appropriate social context, both in and outside the classroom' (OLSET, 1994: 1).

South Africa's Deputy Minister of Education in 2006, Mr. Surty saw the programme as supporting provincial departments in providing in-service teacher development and supporting foundation phase teachers. He praised it for promoting multilingualism and developing the learners' English vocabulary which they need to learn beyond the foundation phase by improving their competence (Surty, 2006). Testimonies from teachers suggesting that the programme has improved their quality of teaching and empowered their professional roles, and that the radio lessons have apparently added value to learners and improved their English language proficiency (Arnott, Mansfield & Mentis, 1993 a; 1993b; Jacobson, 2001).

From the successful development and growth of this programme (from reaching 14 5000 to 1 302 728 learners and counting) interactive radio lessons can be shown to be feasible in developing countries (Naidoo & Potter, 2005). On the basis of widespread acceptance and usage of the programme in South Africa, Nigeria and Sudan there has been an interest in the programmes work internationally particularly in Asia and the rest of Africa (Perraton, Robinson & Creed, 2001). The progress of this programme has led to the need to the validating of the instrument employed to measure the vocabulary of learners receiving. Consequently to this validation, a test manual can be developed.

CHAPTER 3

LITERATURE REVIEW

Introduction

There are two parts to this literature review, the first, Part A reviews test validation literature which incorporates all the stages involved from developing a test to correctly administering it and making appropriate conclusions from the test scores. Part B which follows immediately after focuses on this research's specific construct, vocabulary, and reviews literature defining what vocabulary is, how it can be measured, why it should be measured, why it is important and how the process of learning it can be improved.

PART A

CONSTRUCT VALIDATION

> Validity is an overall evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment (Messick, 1995: 741)

Psychometric and most other tests are 'measuring instruments intended to numerically describe a characteristic under uniform, standardised conditions' (Haladyna, 2004). Some human characteristics are measurable in tangible, quantifiable means; there are those that are not such as constructs measured in the psychology and education fields. Constructs in these fields are often complex and abstract traits which are difficult to quantify or measure such as reading, listening, creative thinking, critical thinking and problem solving which reflect some degree of knowledge or skills (Haladyna, 2004). The processes involved in defining and measuring these constructs is the concern of construct validity. The issues discussed in this section include explaining the role of validation in psychometric tests, highlighting recommended procedures of validating a test and identifying the different components of test validation.

3.1. Defining Test Validation

> "Validity exists in degrees and is a property of test interpretation or use, not of tests" (Haladyna, 2004: 10)

Psychometrics, which is the branch of Psychology concerned with measurement and testing emphasises that tests should have qualities of reliability and validity; Reliability being that the test should be consistent in producing the same scores when administered to the same person all other factors being equal and Validity is the property of the test measuring that which is claimed to measure. Thus, validity is concerned with ensuring that a test measures what it is said to measure and how well it does this so that the inferences made from the test scores are appropriate (Anastasi, 1988; Kubiszyn & Borich, 2003). The process of calculating the reliability and validity of a test instrument is called validation. This process involves calculating the reliability of the test through various, appropriate methods such as Inter-Item and Split-Half reliabilities, depending on the test format (these are other methods of calculating reliability are described in Chapter Four).

Validating a test also involves calculating the validity of the test by using various methods also described in Chapter Four. Some of these validity measures are concurrent, content and construct validity. Another method of validating a test, which can neither be classified as Reliability nor Validity is Item Analysis. The aim of Item Analysis is improving test items by calculating their levels of difficulty and discriminative value which will demonstrate by producing a p-value whether that item is too easy, difficult, or of medium difficulty based on the p-value score produced. Item Discrimination indicates also by a p-value score, how well the particular Item on the test can differentiate the test takers who know the construct being measured and those who do not. Therefore, Item Analysis not only improves the test items but also improves the test because when the test items have good difficulty and discrimination properties the results of the test scores are considered more accurate. The validity of a test score interpretations depends the evidence of reliability, validity and Item analysis which supports that interpretation or use (Haladyna, 1994).

Test Validation can thus be said to be the process of putting a stamp of approval on the test upon gathering sufficient evidence that supports using the test to measure the construct it desires to measure. The Standards for Educational and Psychological Testing stipulates that all psychological tests must have accompanying evidence for validity for all important inferences which are made from the test, meaning, and prior to administering a test and making inferences from it, evidence of its psychometric properties including validity and reliability should be provided (American Psychological Association, 1985). Messick (1984, 1989) has also consistently argued that educational tests must undergo construct validation

prior to use. Therefore, this study's main goal is validating the OLSET Picture Language Vocabulary Level 1 Test by calculating its psychometric properties.

There are three main types of cognitive behaviour constructs in Psychology; these are (i) the construct of ability which refers to intelligence or scholastic aptitude, (ii) the construct of educational achievement referring to the mastering of the learning material indicated by the scores attained in knowledge and skills and (iii) the construct of competence which encompasses the programme of study, similar to educational achievement but also including knowledge acquired through life experiences. This research focuses mainly on the construct of competence, the construct of the English vocabulary, but the test being validated (in other words, the OLSET Picture-Language and Vocabulary level 1 Test) could also be used to assess educational achievement if the programme on which it is based (the English in Action) was an intervention for Level 1 learners to improve their vocabulary and especially if there was pre testing (before the intervention) and there is post testing after the intervention was conducted.

3.2. The Process of Construct Validation

Construct Validation has three sequential steps; Formulation, Explication and Validation this next section will take you, the reader, through the different logical steps that undertaken in validating the construct of vocabulary in the OLSET Picture-Language and Vocabulary Level 1 Test. First, the

Formulation

The first step in testing/measuring is developing the test. This involves naming the theoretical construct to be measured (in this instance, vocabulary), defining the construct (for example, what vocabulary is) and identifying its connectedness with other constructs.

As mentioned in the introduction to this study, the OPLV level 1 Test was originally designed in 1992 and has been amended over the years by different researchers working for the Open Learning Systems Education Trust (OLSET). OLSET has been working in South Africa since 1992. Its aim has been to develop a model for teaching English as a second language in South African primary schools, through the medium of interactive radio and provide in-service training and support to the large number of teachers involved in teaching the programme (Potter & Naidoo, 2006). The programme's growth to scale suggests reasons its acceptance by teachers and learners, as well as its endorsement by educational officials, as it has grown

to scale from providing support to an initial 360 teachers and 14 500 learners under USAID funding in 1993, to an estimated 45 000 teachers and 1 800 000 learners under maximal Department for International Development funding in 2004, and an estimated 32 500 teachers and 1 300 000 learners under funding from the Dutch government by 2007 (Potter & Naidoo, 2006). The final amendment of the OPLVLT 1 before this validation study was undertaken in 2007 by a Research psychologist working in Test Development.

Explication

This stage of the validation creates and identifies the measures of the construct under study, to be on the safe side, Haladyna (1994) suggests that test developers use multiple items of the construct that will cover different aspects of the construct. The OPLVLT uses 36 measures in the form of Multiple Choice Questions. This essentially means that the OPLVLT uses one format of measure for vocabulary, that of word/name recognition. In validating the OPLVLT 1 however, a Metropolitan School Readiness subtest was used as a validating instrument, and this test employed a different format of measure, comprehension. Reasons for employing a singular type measure for the OPLVLT included the fact that most of the learner's to whom the test will be administered are not English speakers (only 4.17% indicated speaking English as a first language) and thereby assuming that their comprehension of long, complex English sentences may be poor; the learner's were also in their first grade of school and 44% of them (according to the SAIRR, 2008) would have not had prior exposure to English before this, other tests measuring vocabulary (such as the Peabody and the WISC) have been shown to be highly reliable and these have mostly used the multiple choice format (MCQ's). For these apparently valid reasons the OPLVLT adopted the MCQ format and consisted of 36 Items.

PART B

VOCABULARY

> "All other things being equal, learners with big vocabularies are more proficient in a wide range of language skills than learners with smaller vocabularies, and there is some evidence to support the view that vocabulary skills play a significant contribution to almost all aspects of L2 proficiency" (Meara, 1996: 37).

There is a growing recognition in the field of language teaching, that vocabulary development is a vital aspect of children's language development (Finocchiaro, 1969). Vocabulary development has been shown to be closely linked to children's progress in school and in their accomplishment in learning to read. Generally, the ability to use English is widely recognised as a prerequisite to success in work, business, and higher education (Dunn & Dunn, 1981). It is essential, therefore, that all children have regular opportunities to learn new words and to incorporate these words into their daily conversations (Wasik, 2006). Vocabulary is increasingly been seen as an important element in language, as a result, second language vocabulary acquisition has recently being receiving increasing attention from researchers, teachers, curriculum designers, theorists and others involved in second language learning (Shen, 2008). This section of the literature review focuses on the importance of vocabulary development, different methods of teaching vocabulary and how methods of assessing vocabulary in the South African context for the learners in Grade One.

3.3. Vocabulary Defined

Vocabulary is defined as the mental store of information about words that include semantic information and syntactic information and word forms (Pinker, 1984). *The American Heritage Dictionary* (2007) defines vocabulary as "the sum of words used by, understood by, or at the command of a particular person or group". The vocabulary of a person can also be defined as the set of words likely to be used by that person when constructing new sentences (http://en.wikipedia.org). Possessing this knowledge of words allows children to accurately label objects and people, learn new concepts, and communicate with others (Wasik, 2006). Vocabulary is often studied in terms of size and breadth (Henriksen, 1999; Qian, 1998, 1999;

Read, 1988, 1989; Shen, 2008; Wesche & Paribakht, 1996), where vocabulary breadth refers to the number of words the meaning of which a learner has at least some superficial knowledge, in other words, the number of words that a person knows and depth of vocabulary knowledge refers to the learner's level knowledge of various aspects of a given word, or how well the learner knows this word (Shen, 2008). Depth of vocabulary knowledge contains components such as pronunciation, spelling, meaning, register, frequency, and morphological, syntactic, and collocation properties which are all structurally and functionally interconnected (Qian, 1998; 1999). There are four identified types of vocabulary, these are; expressive (or productive) vocabulary is concerned with the ability to speak and write a language; receptive vocabulary concerned with listening and reading a language, meaning or oral vocabulary a combination of listening and speaking vocabularies, literate vocabulary combination of reading and writing.

3.4. Knowing a word

Cronbach (1942) divided vocabulary knowledge into two main categories: knowledge of word meaning (generalization, breadth of meaning, and precision of meaning) and levels of accessibility to this knowledge (availability and application). That is, knowing a word is its meaning and being able to use it in appropriate contexts each time you need it. This definition lacks the aspects of lexical knowledge, such as spelling, pronunciation, morpho-syntactic properties, and collocation. Hence, researchers in this field continued to investigate and explain what it means to know a word. Secondary vocabulary researchers have proposed various but complementary frameworks about what it means to know a word. Most agree that lexical knowledge involves degrees of knowledge is not an all-or-nothing phenomenon; they suggest it should be constructed as a continuum, or continua, consisting of several levels and dimensions of knowledge (Shen, 2008).

Richards' well-known vocabulary knowledge framework (1976) identified seven aspects of word knowledge, they are; (i) syntactic behaviour, (ii) associations, (iii) semantic value, (iv) different meanings, (v) underlying form and (vi) derivations). These are simply, knowing the degree of probability of encountering that word in speech or print; knowing the limitations on the use of the word according to variations of function and situation; knowing the syntactic behaviour associated with the word; knowing the underlying form of the word and the derivations that can be made from it; knowing the network of associations between that word

and other words in the language; knowing the semantic value of a word; and knowing many of the different meanings associated with a word, Boers et al. (2004) completely endorses this definition.

Nation (1990) further identified eight types of word knowledge (e.g. form, grammatical pattern, meaning, function, relation with other words), specified for both receptive and productive knowledge. Chapelle (1998) went on to suggest that a trait definition of vocabulary should contain four dimensions: (a) vocabulary size, (b) knowledge of word characteristics, (c) lexicon organization, and (d) processes of lexical access. Henriksen (1999) proposed three separate but related vocabulary dimensions: (a) a "partial-precise knowledge" dimension, (b) a "depth of knowledge" dimension, and (c) a "receptive-productive" dimension. Qian's (2002) recent framework, similar to Chapelle (1998) proposed that vocabulary knowledge comprises four intrinsically connected dimensions: (a) vocabulary size, (b) depth of vocabulary knowledge, (c) lexical organization, and (d) automaticity of receptive–productive knowledge. There is thus a clear consensus that vocabulary knowledge should at least comprise two dimensions, which are vocabulary breadth, or size, and depth, or quality, of vocabulary knowledge (Shen, 2008).

3.5. Vocabulary Acquisition

Vocabulary is acquired in an incremental fashion, so words acquired at the beginning of the learning process are likely to have much more depth than words more recently learned. The more words a learner knows, the more likely it is that he or she will have a greater depth of knowledge for these words (Qian, 2002). Vocabulary acquisition involves the breadth of learning (number of words one can learn), depth of learning (how much a speaker can know about those words). A child is said to 'acquire' a language when that language is the primary language spoken in that child's environment. A child's capacity to learn new words is thus a cognitive ability negotiated between two brain regions. These are the Left frontal lobe of the brain which is responsible for speech production and the Right frontal lobe which is responsible for speech comprehension (Schiefelbusch & Pickar, 1984). The process of acquiring a language undergoes remarkable development during childhood (Gatherole & Baddeley, 1990). The rate at which a child acquires vocabulary is closely related to their general linguistic abilities (Gatherole & Baddeley, 1990) which means that biological factors that influence language acquisition such as the ability to hear are the determining factors that influence language acquisition. Understanding the process of 'learning' a language is

important in order to promote maximal vocabulary development by identifying the component skills involved in acquiring a new item or word of a language. Part of this process has been shown in correlational studies where it was suggested that immediate phonological memory (which is the memory of sound inputs recently heard) is associated with the level of vocabulary in children (Gathercole & Baddeley, 1990).

Learning a new language (also referred to as second language acquisition) on the hand means that the language learnt is not the one primarily used in the child's social context. Children play active roles in their process of learning a language, which involves listening on others as they talk and exposing themselves to situations where others are interacting and participating in that language (Aukrust, 2007). The amount of exposure to rich input spoken in a semantic context and the frequency of a child engaging in a dialogue with a skilled speaker of a language, which is diverse in vocabulary and has a complex discourse, promotes language teaching (Childers & Tomasello, 2002; Holloway, 1994). The more language that children are exposed to and hear, the larger their vocabularies tend to be. Children who already speak another language fluently may improve this process by using their first language as an interpretive resource when trying to figure out the meaning of new words (Aukrust, 2007). The context in which a learner of a new language finds themselves in is important to their learning because vocabulary is best learned in context. Also, because language is seen as social practice, culture becomes core to the process of learning and should be integrated in the learning process. A child needs to be able to produce and comprehend utterances on the basis of social appropriateness as well as grammatical well formedness and referential accuracy in the language that they wish to express themselves in (Schiefelbusch & Pickar, 1984). Failure to develop these language skills impoverishes social interactions grossly (Reisberg, 2001).

3.6. The importance of vocabulary

> Human beings have to be able to communicate in order to be fully functional beings (Crawford, 1993).

The importance of vocabulary in language acquisition goes uncontested (Boers et al., 2004), and it does seem hard to overstate the importance of vocabulary—not only for reading achievement but also for general social and economic success. *The Report of the National Reading Panel* (2000), for example, concluded, "The importance of vocabulary knowledge has long been recognized in the development of reading skills' (p.4). Other studies that have

emphasised the importance of vocabulary particularly with regards to reading, academic success and general success in life are Alderson (2000); Read (2000); Tran (2007); Aukrust (2007); Scheppers (2006); Chan et al (2008); Shen (2008) each of them emphasising the crucial role that vocabulary plays in not only reading but in other aspects of life. Haynes and Baker (1993) also came to the conclusion that the most significant handicap for L2 readers is not lack of reading strategies but insufficient vocabulary in English, mostly because Vocabulary knowledge is correlated with general knowledge or knowledge about the world.

In the past, the size of a person's vocabulary was a very good predictor of that person's general intelligence' (Terman, 1918) but recent literature in this area regards Vocabulary as the best single index of school success and is influenced by exposure and cultural influences (Aukrust, 2007). The motivation for vocabulary instruction is to increase students' speaking or writing vocabulary, improving scores on standardised tests and specific concepts in content areas (McKeown & Curtis, 1987). Possessing a rich vocabulary also increases knowledge about the world and the world that exists beyond what they see in their immediate environment (Blachowicz, Fisher & Ogle, 2006; McKeown & Curtis, 1987). Vocabulary teaching is therefore necessary in developing understanding, speaking, reading and writing abilities within the social and cultural situations. In the years during which children develop as readers and writers, there is an increasingly high relationship among all four aspects of vocabulary—listening, speaking, reading, and writing (Pikulski & Templeton, 2008). Thus, vocabulary also facilitates a gradual perfecting of the language skills needed for everyday communication, improves the learning of other curriculum areas like Science and Maths (Finocchiaro, 1969) and promotes later academic achievement in general (Aukrust, 2007).

According to Nieto (2000) the fluency in English in South Africa is still linked to learners' future economic and social welfare hence the emphasis on the mastery of this language. Researchers have linked vocabulary knowledge to word reading (Torgesen et al, 1997; Carver, 1997; Gottardo et al., 2008; Catts et al., 2008).

3.7. Vocabulary Instruction

> Perhaps the greatest tools we can give students for succeeding, not only in their education but more generally in life, is a large, rich vocabulary and the skills for using those words. Our ability to function in today's complex social and economic worlds is mightily affected by our language skills and word knowledge (Pikulski & Templeton, 2008: 1).

Vocabulary teaching is mainly to facilitate the comprehension of a text that students will be assigned to read to enhance comprehension. Teachers should determine if there are any new words that represent concepts that are critical to understanding the selection *and* which are not adequately defined in context (Pikulski & Templeton, 2008). If there are, then these words should be presented and discussed first, these words would form part of the high frequency words as described by Nation (2001). The English language is said to have the largest vocabulary when compared to other languages; with a total of 171,476 words plus 47,156 obsoletes and 9 500 derivative words. A child of between 6-8 years of age is estimated to know between 6000-7000 words (Watts, 1944) or an average of 8000 root words according to Carey (1978, in Se'ne'chal & Cornell, 1993). The number of words a child knows is determined by exposure to the language and their phonological memory skills (Gathercole & Baddeley, 1990). Nation emphasises teaching of high-frequency words which is a large proportion of running words in spoken and written texts and occur in all kinds of uses of a language (2001). High-frequency words are usually the first 1000 words that a language learner learns. Nation (2001) suggests that prior to the implementation of any language programme instructors should decide on the number of words in the language that they set as their goal, the number of words known by native speakers and the number of words needed to use the language.

3.8. Vocabulary Instruction for young children

The use of pictures is a favourite tool for language teachers; it allows them to integrate culture with vocabulary acquisition and is also facilitative in improving memory (Bush, 2007). Visual representations may precipitate the semantisation process, especially with concrete picturable words such as table where the picture serves as an associate aid to constructing the conceptual networking. In the cases where the word being taught cannot be depicted by a picture Bush (2007) suggests that whatever available appropriate association which can be made with the word, be it an event, behaviour, emotion, sound, should be used to create a cognitive representation of the word being taught. Teachers over the years testify to the effectiveness of using pictures and claim that this teaching method has a significant impact on the learning and retention of vocabulary (Deno, 1968; Kellogg & Howe, 1971; Kopstein & Roshal, 1954 all cited in Bush, 2007).

Young children naturally learn to communicate through listening and speaking (Pikulski & Templeton, 2008). Children who already speak another language fluently may improve this

process by using their first language as an interpretive resource when trying to figure out the meaning of new words (Aukrust, 2007). However, Krashen (1985) argued that second language learners of any language should not be forced to produce any output especially during the early stages of second language development, although he agrees with Childers and Tomasello (2002) that learners must be exposed to large amounts of comprehensible input. Other studies though (i.e. Se'ne'chal & Cornell, 1993), argue that learners of any language learn faster and better when they are encouraged to produce output, with minimal writing, even during the early stages of their second language development. The latter view will be adopted for the purposes of this study. Children learning English as a second language should, according to Tran (2007) also be encouraged to do extensive self-study and to take note of words they do not know and bring them to their teacher for explanation. The context in which a learner of a new language finds themselves in is important to their learning because vocabulary is best learned in context. Also, because language is seen as social practice, culture becomes core to the process of learning and should be integrated in the learning process. Comprehending spoken English is certainly not easy for school learners of English, in part because their first language dominates most of their communication; learning how to listen largely through formal instruction in the classroom and with limited exposure to English outside the context of formal study, ways of making this process better are suggested in the section title 'vocabulary teaching' below (Chang, 2007).

Teachers can improve vocabulary by investing on the knowledge of the learners' native tongue which should be considered an added asset to the teaching of a second language (Finocchiaro, 1969). Vocabulary can also be increased by creating a language-and-word-rich environment which is conducive for learners. In this environment teachers can use various opportunities available for learners to read, hear, use, and talk about new vocabulary. Teachers should also make a point of reinforcing new vocabulary in discussions with students and design instruction to promote students' use of new vocabulary as they speak and write and create an environment that encourages incidental word learning through listening and reading which is important to students' general vocabulary development, providing repeated exposure to the words and providing multiple sources of information (Blachowicz, et al. 2006). Teachers also have the ability and skills to influence children's vocabulary development by providing experiences that expose children to new words (i.e. school trips that will expose the learners to things outside of their environment for example, a trip to the sea or the zoo) as this will help them learn the word meanings and use (Wasik, 2006).Other

ways that could help would be to develop good skills in teaching, being willing to prepare a wealth of instructional material, knowing the broad characteristics of the pupils' language and of English structure, awareness of cultural similarities and differences (Finocchiaro, 1969).

Skills of vocabulary acquisition and reading comprehension are central for second language learners. Teachers in this field can improve on this by providing opportunities to say, write and read new words; by using word walls, word maps, keywords and pictures to help students make connections and by understanding which common sounds in English correspond to the first language of the learners. Providing context, using new words outside of class, connecting examples from a world the learners know, teaching words that will appear across academic areas, teaching comprehension monitoring skills are all more ways through which vocabulary acquisition and reading comprehension can be improved (Swanson & Howerton, 2007). One language of learning and teaching which acknowledges these factors is Interactive Radio Instruction programmes which are used as supplementary methods for teaching. This method is adopted by OLSET in their 'English in Action' programme which supplements other teaching methods for optimal learning of English as a second language and on which the Picture-Vocabulary Test is based. This programme and its aims for improving vocabulary learning will be discussed later.

For a child to communicate successfully they therefore need to be able to produce and comprehend utterances on the basis of social appropriateness as well as grammatical well formedness and referential accuracy (Schiefelbusch & Pickar, 1984) in the language that they wish to express themselves in. A child's understanding, mental representation and memory of objects, persons, sequences of events, and categories of things, persons, and happenings are all involved in what he or she knows and learns about language (Schiefelbusch & Pickar, 1984). Failure to develop these language skills impoverishes social interactions grossly (Reisberg, 2001).

For second language learners of a language there may be challenges with ineffective lexical support in listening comprehension which can be attributed to the lack of automatic processes. Second language learners learn some new element of a language, and must pay attention and think about it which takes some time to be eventually incorporated and even a longer time to use (Buck, 2001). These learners thus need more time to practice lexical items to automatise processing.

In order to make the transition to communicating through reading and writing, they need a large meaning vocabulary and effective decoding skills. There is an abundance of research evidence to show that an effective decoding strategy allows students not only to identify printed words accurately but to do so rapidly and automatically (Pikulski and Chard, 2005). The rate at which a child acquires vocabulary is closely related to their general linguistic abilities (Gatherole & Baddeley, 1990) which means that biological factors that influence language acquisition such as the ability to hear, the ability to speak and the ruling out of possible impeding speech disorders are the determining factors that influence language acquisition.

"The single most important activity for building the knowledge required for eventual success in reading is reading aloud to children" (Pikulski & Templeton, 2008: 3). One way in which reading aloud to children can be expected to be beneficial is to increase their language and vocabulary skills. Indeed there is research to support this position (Elley, 1989; Leong and Pikulski, 1990; Robbins and Ehri, 1994). Also to be considered in foreign vocabulary teaching is that it must be presented in a cultural context if it is to be usable (Spinelli & Siskin, 1992). Learners must be encouraged to blend in the new words and concepts into their already existing vocabulary and culture. Croll (1971) also adds that, learners remember words which resemble something already existing in their environment as this makes the process of integration quicker. Teachers can thus improve vocabulary by investing on the knowledge of the learners' native tongue which should be considered an added asset to the teaching of a second language (Finocchiaro, 1969).

Vocabulary can also be increased by creating a language-and-word-rich environment which is conducive for learners. In this environment teachers can use various opportunities available for learners to read, hear, use, and talk about new vocabulary. Providing context, using new words outside of class, connecting examples from a world the learners know, teaching words that will appear across academic areas, teaching comprehension monitoring skills are all more ways through which vocabulary acquisition and reading comprehension can be improved (Swanson, 2007). One language of learning and teaching which acknowledges these factors is Interactive Radio Instruction programmes which are used as supplementary methods for teaching. This method is adopted by OLSET in their 'English in Action' programme which supplements other teaching methods for optimal learning of English as a second language and on which the Picture-Vocabulary Test is based. This programme and its

aims for improving vocabulary learning will be discussed in more detail in the final research report.

Learners who are being taught in a second or third language require unique skills to be taught effectively. L2 learners need only partial knowledge of a word in comprehension, although more lexical knowledge is obviously desirable in many situations (Qian, 2002). However, in people whose first language is not English it is often considered sufficient to measure the communicative language ability of English (Taylor, Kroch, & Ringe, 2000).

3.9. Vocabulary Assessment

> An Assessment is a machine for reasoning about what students know, can do, or have accomplished, based on a handful of things they say, do, or make in particular settings (Mislevy et al., 2003: 4).

Assessment enables the monitoring of progress (Nation, 2006). Vocabulary assessment is the measurement of students' knowledge of words and their meanings. Vocabulary was first assessed when intelligence was first measured by Binet and Thurstone (Nation, 2006) where students had to define or explain words. According to Mislevy et al. (2003), an assessment can be summarised as the machine for reasoning about what students "know, can do, or have accomplished, based on a handful of things they say, do, or make in particular settings" (p. 4)

These tests were later adapted in such a way that they could be administered to larger groups and that is when standardised, multiple-choice versions of assessments were born, during World War 1 (Resnick & Resnick, 1977).

The purpose of vocabulary assessment is to "monitor the learner's progress in vocabulary learning and to assess how adequate their vocabulary knowledge is to meet their communication needs" (Read, 2000: 2). Pearson, Hiebert and Kamil (2007) argue however, that vocabulary assessment is undernourished in its theoretical and practical aspects. They believe that vocabulary assessment has been driven by convenience and psychometric standards as opposed to a clear conceptualisation of its nature and relation to other aspects such as comprehension. They feel that vocabulary measures do not take into account the global measures of comprehension in their measurement of words. They insist that the instruments used to measure vocabulary should capture the relationship between the words themselves and how these will influence comprehension of the language learned. The use of

the ‘school language and listening subtest of the Metropolitan Readiness Test is meant to address this gap.

Also central to vocabulary assessment, or any testing for that matter, is context. When children are asked by ‘strange people to perform strange actions or answer strange questions in strange rooms or situations’ the result is usually distorted from the reality (Holland, Goodman, & Walkei, 1994). Therefore the nature of that assessment is not a reflection of the natural context in which the children learn and this means that an array of psychological issues are at play such as anxiety, these may influence the performance of those test takers. Therefore the context in which learners are tested should resemble to as close a degree as possible their natural context. Essentially, paramount attention should be paid to the social context in which the assessment of the language occurs (Chang, 2007).

> Depending on the task, a person could perform adequately with relatively imprecise knowledge. In other situations, a much finer notion of the word’s meaning might be required” Mezynski (1983: 265).

Measures of vocabulary knowledge are important predictors of a variety of indices of linguistic ability (Anderson and Freebody, 1981: 77), and results from vocabulary testing can have a variety of goals such as placements in language programmes, diagnosis of strengths and weaknesses, encouragement of learning, monitoring progress and measuring profiency as discussed in Chapter 1 (Nation, 2006). Vocabulary must be assessed with standardized, norm-referenced tests (Ukrainetz, 2002) as vocabulary test results provide useful information on how vocabularies develop, are valid indicators of language ability and indicate whether learners have acquired the words they were taught (i.e. achievement testing). Vocabulary test scores also assist in detecting whether there are gaps in the vocabulary of learners and inform decisions on whether to place students in appropriate language class levels (i.e. placement testing) (Boers et al., 2004).

CHAPTER 4

RESEARCH DESIGN AND METHODOLOGY

This chapter discusses the methods undertaken to fulfil the aim of this study which was to validate the OLSET Picture-Language Vocabulary Level-1 Test (OPLVLT) by providing evidence of its psychometric properties including validity, reliability and Item analysis.

4.1. Research Questions Guiding the Study

In order to answer each of the Research Questions set out in this study, the following procedures were followed;

Research Question 1: Is the OPLV Level 1 Test a reliable measuring instrument which produces consistent and accurate scores?

This question will provide evidence that the OLSET Picture-Language Vocabulary Level 1 Test produces consistent and accurate scores on all test takers and thereby making it a reliable testing instrument.

***Hypothesis One*:** The OLSET Picture-Language Vocabulary Level 1 Test is a reliable testing instrument producing consistent results each time it is used.

The overall reliability of the instrument was measured by the Internal consistency's index of Cronbach alpha reporting standardised scores.

Any item which compromised the reliability property of the Test (by decreasing it) was to be deleted from the Test if so doing was considered beneficial for the overall reliability of the Test.

Research Question 2: Does the OPLV Level 1 Test measure vocabulary? That is, is the Test a valid measure of the construct, vocabulary, as it claims to be?

***Hypothesis Two*:** The knowledge measured by the OLSET Picture-Language Vocabulary Level 1 Test is a reflection of the test-takers vocabulary knowledge, as indicated by strong, positive correlations with the Peabody Picture-Vocabulary Test, which has been demonstrated internationally to also measure vocabulary. .

***Hypothesis Three*:** There are not strong, positive correlations between OPLVL Test and the Metropolitan Readiness School Language and Listening subtest, which measures a more general language construct relating to school readiness.

The validity of the test was calculated by using two other tests postulating to measure vocabulary. These were the Metropolitan School Readiness School Language and Listening Subtest and the Peabody Picture-Vocabulary Test. These tests were selected based on their previous cross-cultural usage as measures of language at the foundation phase in primary school, and also their psychometric properties of reliability and validity in relation to the samples on which they were standardised. The tests are described in more detail in section 4.3. Below.

In terms of hypotheses two and three, the OLSET Picture-Language Vocabulary Level 1 Test was expected to correlate with the PPVT if it measured the construct 'vocabulary'. A strong correlation between the OPLVLT with the PPVT would indicate that the OPLVLT is a good measure of vocabulary.

Low correlations between the PPVT and the OPLVLT may suggest that the OPLVLT does not measure the construct variable 'vocabulary' or (ii) that the theoretical network which generated the hypothesis was incorrect or (iii) that the experimental design failed to test the hypothesis properly (Cronbach & Meehl, 1955). However, it would be inadequate to report construct validity in the form of a single simple coefficient, an additional measure (the MRT) which is also a group test constructed for work with young children at the foundation levels in primary school will also be utilised so to substantiate the integration of the data into a proper interpretation of the construct under study (i.e. vocabulary).

Thus, the correlation between the average scores on the OPLVLT with the average scores on the Metropolitan Readiness School Language and Listening subtest and the average scores with the Peabody Picture-Vocabulary Test was calculated by Pearson's Product-Moment correlation. An overall Cronbach's alpha on the scores of the three tests was also calculated as an overall index of consistency of the constructs measured by the three instruments.

Research Question 3: Are there statistically significant differences between the scores of male and female learners on the test?

Hypothesis Four: Vocabulary literature indicates that female learners fair better in language

tests than male learners (Westerlund & Lagerberg, 2008) therefore they would be expected to attain higher scores in the OPLVLT than their male counterparts in this validation study.

Research Question 4: Are there significant differences between the scores of learners who speak English as a first language and those who speak other languages as first language on the test?

Hypothesis Five: Learners who speak English as a first language were expected to fair better on the OPLVLT than the learners who spoke another language as a first language.

Research Question 5: Does age influence the scores on the OPLVLT in a statistically significant manner?

***Hypothesis Six*:** Previous studies have shown that the amount of exposure to rich language input spoken in a semantic context enhances language learning and that the frequency of a child engaging in a dialogue with a skilled speaker of a language, which is diverse in vocabulary and has a complex discourse, promotes language learning (Childers & Tomasello, 2002; Holloway, 1994). On the basis of this background, older learners were expected to have bigger vocabulary sizes and score higher on the OPLVLT test than younger learners because the more language which children are exposed to and hear, the larger their vocabularies tend to be.

4.2. Research Design

The design of this study was non-experimental and was designed for validation purposes. The OPLVLT test is currently used as part of the 'English in Action' English Language Radio learning programme which was designed to improve the vocabulary of learners in the first three grades of primary school. The programme has been designed and implemented by OLSET.

The OPLVLT has previously been used to estimate the vocabulary of the learners prior to the implementation of the 'English in Action' programme and also post the implementation to measure the success of the programme. The validation of the OLSET Test would enable its use across South Africa with greater confidence, if it can be shown that the instrument is a reliable and valid measurement of vocabulary for use in schools and other vocational and education settings.

4.3. Instruments used in this Study

4.3.1. Metropolitan Readiness Test Level 1 (1986)

The School Language and Listening group subtest (Test five) of the Metropolitan Readiness Test Level 1 Test measures the ability to use English in the school context for learners between 4-7 years of age. The School Language and Listening subtest has 15 items, and requires approximately 14 minutes for administration.

The test requires that learners mark underneath the picture which they consider to best correspond with a statement provided orally by the test administrator. An example of an instruction might be 'mark under the picture that shows no cars', the test-takers will have some time to look through either 3 or 4 pictures and then mark under the picture that shows the picture without any cars in it.

This test (i.e. the MRT) was originally validated on three groups at different times of the year in four districts of the United States. The first group administration, on fourteen thousand Kindergarten and Grade One learners from 180 school districts yielded a Kuder-Richardson Formula 20 (KR-20) reliability coefficient of 0.74. The second group, (on 7000 Kindergarten from 140 school districts) had a KR-20 of 0.66. The third group, (on 900 pre-kindergarten and kindergarten learners from 290 school districts) yielded a KR-20 of 0.73.

Most of the schools in the United States validation sample were public schools (>60%), the learners were mostly white (>70%, followed by black (19%), Spanish (7%) and Non-specified other (3%), the Socio-Economic Status (SES) of the learner's family was in the middle class range according to the SES index. The test-reliability of this subscale was 0.82.

Predictive validity coefficients for this subtest are reported as 0.34 for reading and 0.40 for language for the 6th Edition of the Metropolitan Achievement Test and 0.48 for reading and 0.70 for listening for the 6th Edition. The manual on the test also reports performance on tasks relative to Grade norms.

See Appendix C for the test items, administration manual and scoring information.

4.3.2. The Peabody Picture Vocabulary Test-Revised (1981)

The PPVT-R measures the test-takers' receptive vocabulary. The test comprises of two forms, L and M, each with 75 items with the first 5 items as practice items. Form L was used

for this study. The Peabody Picture Vocabulary Test was designed for children from age two and a half years to 16 years of age. Each item of the test consists of a set of four pictures, in Multiple Choice format where only one picture is correct. The task is to select the picture considered to illustrate the meaning of a stimulus word presented orally by the examiner. Results from this sample were used as age-referenced and grade-referenced norms.

Form L of PPVT-R was validated on 4 200 American children between two and half years and sixteen years of age selected on the basis of sex, age, geographic area, occupational background, racial-ethnic and urban-rural population distribution. The test was validated against a number of other vocabulary tests such PPVT-IQS, the WISC-R IQS, Stanford-Binet IQS, Mc Carthy Scales of children's Abilities, Wechsler Full Scale IQS, Stanford-Binet IQS, and McCarthy Scales of children's abilities GCI's. There was no predictive validity or long-term temporal stability.

Reported reliability coefficients of the PPVT-R are split-half coefficients ranging from 0.67 to 0.84, and alternate-forms reliability coefficients ranging from 0.74 to 0.83.

Scoring the PPVT-R is rapid and objective and is accomplished largely while the test is being administered by ticking on the appropriate box indicating whether the option picked by the test taker is correct. The test is administered from a basal level through to the ceiling level (two errors in eight words) is reached. The number of items correct is then calculated. Testing is estimated to take between 10 and 20 minutes for 35 items and is administered individually to each learner.

Data produced by this test is interval because the test yields a numerical score based on items of ascending order of level of difficulty. An example of data produced by this test is;

Child's Name: Vusi

Ceiling Item: 27, Minus Errors: 8,

Raw Score: 19.

The ceiling and basal items serve to illustrate to the test administrator the cut-off points for the test taker in terms of performance. The raw score is then used to report the performance of each child relative to age norms.

For the purposes of validating the OPLVLT, the first 35 items of the test were included in the range of items to be administered individually to learners in Grade One.

See Appendix D for the test, an example of a scoring sheet and reliability information of the test.

4.3.3. The OLSET Picture-Language Vocabulary Level 1

This instrument has been constructed as a group test which consists of 36 items, made up of 20 nouns, 7 verbs, 2 pronouns, 5 adjectives and 2 prepositions. The test is a Multiple Choice format similar to the other instruments used. The OPLVLT is relatively easy to administer and is estimated at requiring 15 minutes to administer. An audiotape of instructions is utilised as opposed to the administrator speaking is used, this was done to avoid discrepancies on the learners' performance due to varying pronunciations by the test administrator and also to enable standardised administration in group or classroom situations in which the English language competence of the children and their teachers is not well developed.

The test was first constructed in the early 1990's. A basic content validation by English language teaching experts was performed (Arnott, Mentis and Mansfield, 1993). Otherwise, no validity and reliability information can be reported, despite the use of the test in a number of studies (eg Jacobson, 2001).

4.4. Sample

The sample were drawn from Grade One classrooms in three primary schools on the East Rand, namely Welizibuko Primary School (which uses English as the language of learning and teaching) in Kliptown, Soweto (n=43); Paul Mosaka Primary School (which uses Sesotho as the language of learning and teaching) in Pimville, Soweto (n=43); and Bree Primary School (which English language of learning and teaching) in Mayfair (n=43) Primary Schools (See Table 4.1. below).

Table 4.1.: Summary of School Frequencies

Name of School	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
Bree	43	35.83%	43	35.83%
Paul Mosaka	43	35.83%	86	71.67%
Welizibuko	34	28.33%	120	100%

The descriptive statistics for gender and age in the sample were as follows:

Gender

The total number of learners who participated was 120. 37 were boys (30.83%) and 83 were girls (69.17%), as reflected in Table 4.2. below.

Table 4.2. Gender Frequencies

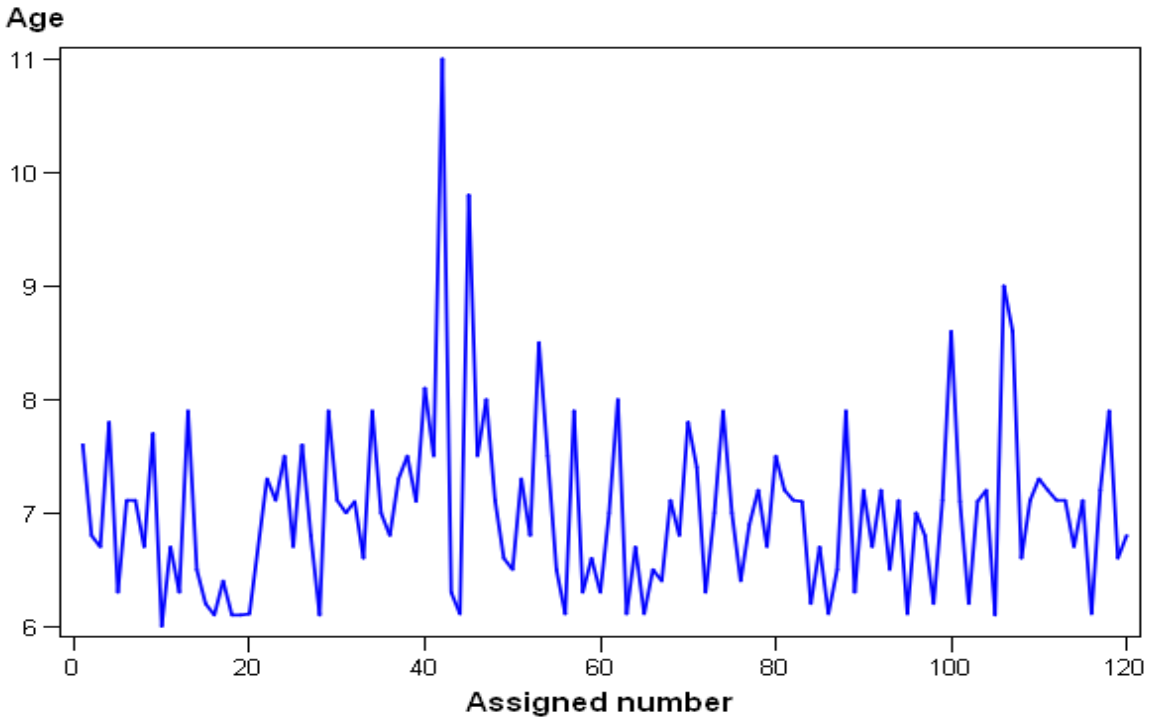
Gender	Frequency	Percentage	Cumulative Frequency	Cumulative Percentage
Boys	37	30.83%	37	30.83%
Girls	83	69.17%	120	100%

The learners who participated were selected randomly from their class lists using a stratified random sampling method from their class lists. The researcher obtained the class list and selected every third name on the list to the sample. The learners who had been selected were then moved to a separate classroom where testing was conducted. It is thus likely that the imbalance of gender in the final sample was an artefact of the sampling procedures used, as opposed to a random influence on the data.

Age

The learners' mean age was 7.02 years old. The youngest learner who participated was 6.0 years old and the oldest was 11.0 years old (See Graph on the next page).

Graph 4.1. Distribution of age of learners



Home Languages

The learners in the sample spoke various home languages, with 37 (30%) of the 120 learners having indicated that they speak Sesotho as a first language, 35 (29. 1%) of the learners speak isiXhosa as a first language, Fifteen (12. 50%) speak isiZulu as a first language, Fourteen (11.66%) speak Setswana and Five (4.16%) of the learners reported to be speaking English as a first language (See Table 4.3. on the next page).

Table 4.3.. Distribution of learners' home languages.

Language	Frequency	Percent %	Cum Frequency	Cum %
English	5	4.17	5	4.17
Malawian	1	0.83	6	5.00
Shangaan	4	3.33	10	8.33
Somali	2	1.67	12	10.00
Urdu	7	5.83	19	15.83
isiXhosa	35	29.17	54	45.00
isiZulu	15	12.50	69	57.50
Sesotho	37	30.83	106	88.33
Setswana	14	11.67	120	100.00

Due to the fact that the study was about validating the OPLVLT which is a measure of the English Vocabulary, all learners were tested in English only.

4.5. Data Collection Procedure

OLSET approached three schools to participate in the validation study. These schools were identified by the Programme Regional Co-ordinators from a range of possible schools. They were chosen on the basis that each of the schools had been consistently implementing their 'English In Action' Radio Learning programme over a number of years, and were thus likely to have sufficient command of English to form a standardisation sample.

OLSET then employed and trained 10 third-year Psychology students from the University of the Witwatersrand to work with the researcher. Each of these students had previously taken courses in research methodology and psychometrics. The students were then trained in test administration by a registered Research Psychologist who is also a PhD student at the University.

Prior to data collection, application was made to the Provincial Department of Education to use these schools for this research. After permission was granted, arrangements were then made with the schools to schedule dates and times for test administration. 45 minutes was requested from a school day and the class teachers were involved and assisted in distributing the workbooks and briefing the learners.

The two group tests (the OPLVLT and the Metropolitan Readiness School Language and Listening subtest) were then administered to all the children. The Peabody Picture-Vocabulary Test was then subsequently administered to all the learners in the sample on an individual level.

OLSET then kept all the test protocols and answer books with the raw scores and made these available to the researcher for the purposes of this study. After the data had been coded, the learners' names and identifying information were then removed from the data set.

4.6. Ethical Considerations

Ethical clearance was obtained by the Open Learning Systems Education Trust through the Gauteng Department of Education (relevant documents are attached). The ethical considerations included obtaining Informed consent from the School Principals and teachers. Assent was not obtained from the learners as the data collected was within the framework of their curriculum and the nature and process of collecting the data posed low-risk (no foreseen physical, psychological, social or economic risk was identified for the learners) and high benefit (test scores would assist in identifying knowledge gaps and intervention design). Learners were informed of their voluntary choice to participate in the study and they informed of their option to withdraw from the study at any point. I thus obtained permission from OLSET to use their data for the purposes of validating the OPLVLT.

4.7. Summary Statistics

In terms of the analysis, all 120 Grade One learners' scores on the three vocabulary tests (the PPVT, MRT and the OPLVLT) were utilised for the purposes of validating the OLSET Picture Language and Vocabulary Level 1 Test. Prior to calculating reliability and validity measures for the OLSET vocabulary test, summary statistics were computed for all test variables for the standardisation sample tested.

The biographical variables were nominal and frequencies were calculated to describe the sample in terms of age, home language, school, gender (as indicated above in 4.3.). Average (mean) scores and standard deviations were then calculated for the different tests, according to gender and school. This part of analysis served the purpose of describing the sample and the dataset. The table below tabulates the average scores of all 120 learners in the three tests.

Table 4.4. Means of scores attained on the tests

Variable	N	Mean	Std dev	Min	Max
Metropolitan	120	8.63	3.65	0	16
Peabody	120	21.90	5.67	9	35
OPLVL	120	33.08	3.13	15	36

The mean scores for boys in the three different tests were; 8.56 (53. 50%) for the Metropolitan School Readiness Subtest, 22. 97 (65.62%) for the Peabody Picture and Language Test and 32. 75 (90.97%) for the OPLVLT. The correlation between the OPLVLT and the Metropolitan Readiness School Language and Listening subtest was 0.38 and 0.31 for the OPLVLT and the PPVT-R. Female learners had higher averages than those attained by boys in the MRT subtest and the OPLVLT and a lower average score on the PPVT-R. On the MRT subtest girls attained 8.66 (54. 12%) as an average score, 21. 42 (61. 20%) score on the Peabody and 33. 18 (92. 16%) average score on the OPLVLT. Gender was thus identified as a potential variable as indicated in Research Three, and the statistical significance of this variable is calculated in Chapter Five.

Seventy-two (72%) of the learners scored above 50% on the Metropolitan subtest, three (3) of the learners scored 0 for the Metropolitan subtest, and 5 of the learners scored 100% on the same test. Most of the learners had scores between 7 and 9 which are in the range of 43.75% and 56. 25%. The lowest score attained in the Metropolitan subtest by boys was 2 whereas it was 0 for the girls. 56. 7% boys scored above 50% and 60% of girls scored above 50%.

More than Seventy-five (75. 83%) of the learners scored 50% (17.5) or more on the Peabody. The lowest score attained by boys in this test was 12 and for girls it was 9. Each gender had 1 learner scoring 100% for the test.

Eight female learners scored 100% on the OPLVLT and three males scored 100%. 90% of the learners scored above 83%. Only one learner scored below 50% in this test, this learner was female. The overall average score on this test 91. 1%.

The scores attained by the learners in the OPLVLT test are graphically displayed below;

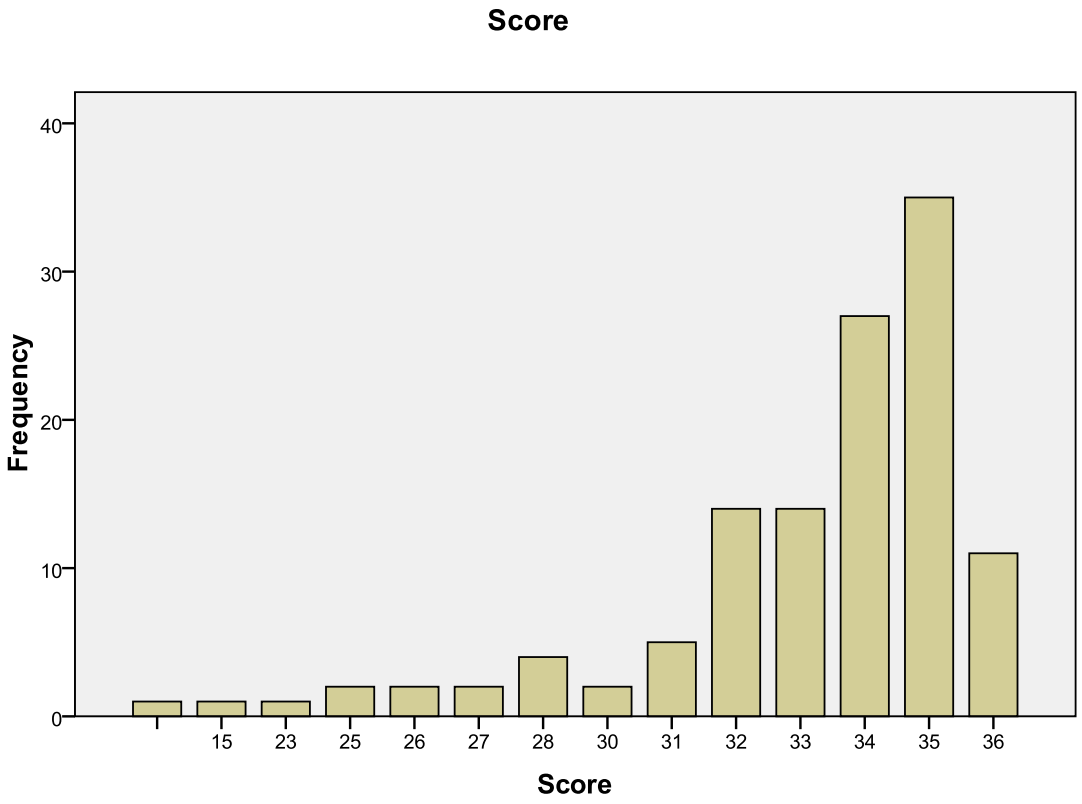


Figure 4.1 Frequency of scores in the OPLVLT

The scores attained by the learners on this test were high, as the skewness to the right demonstrates. Further analyses will expand on the implications of this result.

4.8. Reliability Analyses

The reliability of an instrument or measure refers to the consistency with which the scores measure a certain construct. This is normally estimated from the correlation of an item, scale, or instrument with a hypothetical one which truly measures what it is supposed to (Foxcroft & Roodt, 2001).

In estimating reliability, Internal consistency was estimated by the Cronbach alpha reliability coefficient was calculated, this coefficient produces scores in the range between 0.00 and +1.00. What Cronbach's alpha essentially indicates is the average of all the possible split-half estimates of the instrument's reliability from a particular sample. These split-half estimates are the product of all the possible split-half sets, meaning that all possible items have been paired into two splits of the instrument and the overall alpha thus estimates this reliability coefficient. SAS was used to perform this analysis and this software does the random subsets

of items and computes the resulting correlations. The figure below illustrates the different split-half estimates and how the overall Cronbach coefficient is derived.

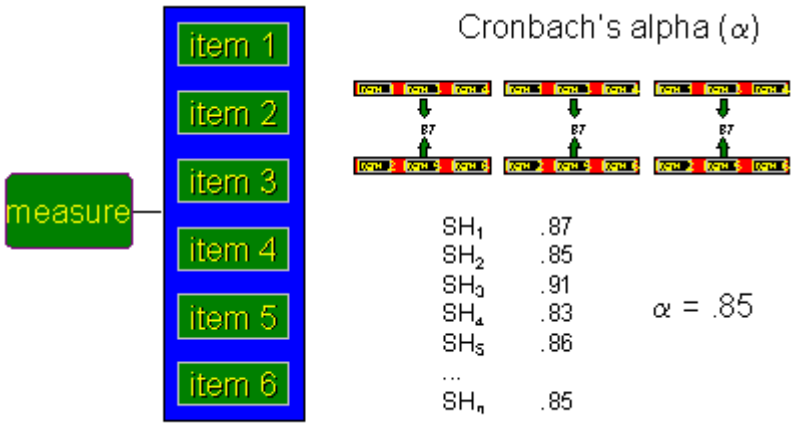


Figure 4.2: Cronbach's alpha illustration

Because the OPLVLT is an achievement test and certain decisions are made from the score yielded, the desirable reliability coefficient should ideally be 0.08 and above (Cronbach, 1955).

Alpha if deleted. SAS computed "Cronbach's Alpha if Item Deleted," which gives the researcher an option to drop items with high coefficients in this column as another way to improve the alpha level, suggesting that the item might not measure the construct it is said to measure (thus, the particular item has low construct validity).

The item-total correlation. This is the Pearson correlation of the item with the total of scores on all other items. A low item-total correlation means the item is little correlated with the overall scale and the researcher should consider dropping it.

Negative alphas. A negative Cronbach's alpha indicates inconsistent coding or a mixture of items measuring different dimensions, leading to negative inter-item correlations.

4.9. Validity Analyses

The validity of a measure refers to whether a test measures what it claims to measure and the extent to which it measures it for the specific purpose that it is set out to measure it (Foxcroft & Roodt, 2001). A validity coefficient thus indicates the extent that inferences made from a test score are appropriate, meaningful and useful (AERA, 1985). For the purposes of this study the Construct Validity coefficient (as indicated by the Pearson Product-Moment

correlation coefficient) will be measured and used as an estimate of the relationship between an instrument and its abstract, theoretical, intangible variables (Cronbach, 1988). In this study, the coefficient of the relationship between the OLSET Picture-Language Vocabulary Test and both the Metropolitan Readiness School Language and Listening Subtest and the Peabody Picture Vocabulary Test will be used to estimate the inferred, underlying characteristics common to these tests to be established. In this way it is hoped to substantiate the vocabulary construct underpinning the OLSET Picture-Language Vocabulary Level 1 Test, in terms of one external referent (the Peabody Picture Vocabulary test). The relationship with another commonly used group test of language (the Metropolitan Readiness School Language and Listening subtest) will also be demonstrated. This will be done to validate the existence of the vocabulary construct through establishing a network of interlocking suppositions, from which can be derived an underlying theory about the construct (AERA, 1985).

The construct of any concept is described as its abstract, intellectual property which manifests itself in observable and measurable phenomena (Reber, 1985; Rosenthal & Rosnow, 1991). *Construct validation* 'is involved whenever a test is to be interpreted as a measure of some attribute or quality which is not "operationally defined."' (Cronbach & Meehl, 1955). Construct validation is mandatory in studies where the phenomena investigated has no criterion or universe of content which is accepted as entirely adequate to define the quality to be measured (Cronbach and Meehl, 1955). Construct validity is ordinarily studied when the tester has no definite criterion measure of the quality with which he/she is concerned, and must use indirect measures, being able to determine the psychological constructs which account for a test is desirable for almost any test (Cronbach and Meehl, 1955). Here the trait or quality underlying the test is of central importance, rather than either the test behaviour or the scores on the criteria (Cronbach and Meehl, 1955). Construct validity would be involved in answering such questions as: To what extent is this test culture-free? Does this test of "interpretation of data" measure vocabulary?

4.10. Group Differences Analysis

Non-parametric ANOVA: Differences between male and female learners' scores

A Non-parametric Analysis Of Variance (i.e. Kruskal Wallis ANOVA) was calculated to establish any statistical differences between the performance of male learners and female learners on the OPLVLT.

The Degrees of Freedom, Sum of Squares, Mean Squares, F-statistic, and probability value indicating the statistical significance at Alpha 5% are reported to indicate any differences between learners in the different language groups. This statistical technique tests the null hypothesis that the means to the groups are equal if a randomised design is involved because adjusted and unadjusted population means are the same.

ANOVA (Parametric Analysis of Variance) was used to estimate whether non-parametric and parametric analyses yielded different results. This step was taken to establish whether ANCOVA (described in the following section) could legitimately be used.

ANCOVA: Age as a covariate

The Analysis of COVAriance (ANCOVA) is a technique that has aspects of ANOVA and regression analysis and increases the precision of comparisons between groups by accounting to variation on important prognostic variables and adjusting comparisons between groups for imbalances in important prognostic variables between these groups (Huitema, 1980). The model employed in ANCOVA not only treats variance between groups as would ANOVA or variance accounted for by regression as systematic as a regression analysis would, ANCOVA treats both simultaneously (Huitema, 1980). ANCOVA is used to test the null hypothesis that two or more sample means were obtained from populations with the same mean. ANCOVA has greater power and a reduced bias due to differences between groups that exist before experimental treatments (Huitema, 1980). Similar to ANOVA, the ANCOVA F-test concerns the null hypothesis that two or more population means are equal if a randomised design is involved because adjusted and unadjusted population means are the same (Iversen & Norpoth, 1987).

For the purposes of this study was to calculate if there is any statistical significance on the score of the learner based on their age. Age was calculated as a covariate, with gender being the categorical/class variable and the results include the F-statistical, the p-value which indicates significance, the R-square and Type III adjusted mean.

In summary, four analyses were conducted to validate the OPLVLT instrument; these were reliability, validity, Non-parametric ANOVA and ANCOVA analyses. Reliability was measured by Cronbach alpha for all three instruments and Construct validity was measured and reported through Cronbach alpha and indicating the correlational strength of the OPLVLT with the Peabody and Metropolitan instruments which were also administered. Group differences in the learners' scores were measured through two variables; gender and age. Gender as a variable was calculated through analysis of variance and age was calculated through the analysis of covariance.

Underlying Validity and Reliability assumptions made about the data were that there was no selection bias, no information bias, and comparability of groups when comparing the effects of the exposure to the EiA programme (no confounding). These assumptions were confirmed by the researcher to have been met during the process of data collection.

The following assumptions were made for the Non-parametric ANOVA about the data. The first assumption was that there was normality in the data; that is, at least 95% of the sample should lie within two standard deviations of the mean in order to conform to a normal curve this characteristic of the data is needed for statistical inference. Normality of the data was examined by through the Shapiro-Wilk and Kolmogov-Smirnov tests and the normal probability plot. The OLSET scores did not conform to this assumption and the data was heavily skewed to the right. The results are described in the Findings section in Chapter 5.

The second assumption for these analyses was that there was homogeneity of variables which means the population from which the data is sampled should have the same variance; Levene's test and its significance provide evidence for homogeneity of data. The results for the Levene's test are discussed in Chapter 5.

The third assumption made was that data was independent and that there was randomised, independent sampling methods used. These assumptions were met as indicated in the next chapter.

The fourth assumption was that there is equality of variance within and between groups for the data compared and Tukey's test indicates this; this test shows if there is no correlation between error terms and the Independent variables (Dean & Illowsky, 2008).

The fifth assumption made in the study was that the means are additive.

The decision guiding the choice of statistical method to use was then guided by the decision tree on, figure 5.1 on page 69.

Over and above the similar ANOVA assumptions, ANCOVA assumptions were that there is independence of the covariate (i.e. age), that there is homogeneity of the regression slopes through the F-test and that there is a linear relationship between the covariate and the dependent variable (i.e. test score) by checking that the p-value is not significant at the level of significance stipulated by the researcher. R-square reported the amount of variation in the test score that the model (i.e. age) could account for. An F score more than 2 or less than negative 2 is considered large and would usually be significant at 1% or 5%. An F closer to 1 often indicates that there the sample means come from populations with equal means and also indicates whether the model as a whole accounted for a significant proportion of the score on the test and p-value indicated overall significance of the effect of the variable (Iversen & Norpoth, 1987; Tredoux & Durheim, 2002). Chapter Five presents the results of these analyses.

CHAPTER 5

FINDINGS

This chapter reports the findings of the study and presents the results of the analyses discussed in Chapter 4. A number of analyses were undertaken to validate the OLSET Language Picture-Vocabulary Level 1 Test; these were Cronbach's alpha as an index of internal consistency, Pearson's Product-Moment correlation as an index of construct validity, Non-parametric ANOVA, parametric ANOVA, and ANCOVA.

The reliability analyses examined the ability of the three test instruments' abilities to measure consistently each time by reporting Cronbach's alpha for all possible combinations of the items of the scale. The validity analyses examined the construct validity of the instrument, by examining the relationship between the instrument and two other tests. One was another test of vocabulary, and the other a scholastic test measuring school language and listening skills, as a measure of the foundations for learning in primary school. The Analysis of variance examined the relationship between the gender of the learner and the score they attained on the test.The Analysis of covariance measured the strength of the covariate, age, and its influence if any, on the scores attained on the instrument.

Table 5.1 Research questions, statistical techniques and variable measured

Research Question	**Statistical Technique/s Used**	**Variable Measured**
1. Is the OPLV Level 1 Test a reliable measuring instrument which produces consistent and accurate scores?	1.1. *Internal consistency*; Cronbach's Alpha. 1.2. *Split-half Reliability*; Cronbach's Alpha, Pearson Product-Moment Correlation and Spearman-Brown Coefficient.	All 36 Items on the OPLVL Test
2. *Does the OPLV Level 1 Test measure vocabulary?*	Pearson' Product-Moment Correlation between the OPLVLT and the two tests acting as criteria for validation. Overall Cronbach's alpha as an index of the consistency between all three tests.	Scores on the OPLVLT, the Peabody Picture Vocabulary Test, and the Metropolitan Readiness School Language and Listening subtest
3. *Are there statistically significant differences between the scores of male and female learners on the test?*	Descriptive statistics; Means and Standard deviations, Non-parametric Analysis of Variance on Gender as a variable. Parametric Analysis of Variance with Gender as a variable.	Dependent variable: Score on the OPLVLT Independent/Class Variable: Gender
4. *Are there statistically significant differences between the scores of learners who spoke English as a first with those who did not?*	Unable to calculate as the data collected had only 5 learners who reported to speaking English as a first language with the remaining 115 speaking other languages either than English.	N/A
5. *Are there any **differences** of statistical significance between scores of learners of different ages?*	Descriptive statistics; Means and Standard deviations, Analysis of Covariance on the variable Age	Dependent Variable: score on the OPLVLT Continuous Variable: Age Class Variable: Gender

5.1. Reliability

Any test of measurement has three major sources of error: factors in the test itself, factors in the learners taking the test, and factors in scoring the test. Reliability measures assist in estimating these errors and are thus of central importance in any new test (Griffin, 2005). When the content of the test is too dissimilar reliability of the test is compromised, ideally the content of the test is expected to be similar thus to create a relatively highly reliable test.

Another source of error particularly on the OPLVLT is the inclusion of distracter items (wrong options), partially correct distracters, and more than one correct answer on the multiple choice options because when skill and domain become complex more errors are more likely to occur (Haladyna, 2004).

Learners themselves are not always consistent and this inconsistency introduces error into the testing process. Other factors such as learners' attitudes, anxiety levels, health, and sleep may affect the performance of the learners and thus affect their test-taking consistency (Haladyna, 2004). In scoring the tests there is also potential for errors to occur, however, for the OPLVLT scoring error was minimal because it was mechanical (choosing the marked response). The reliability analyses were calculated to identify and estimate the reliability of the test instrument (i.e. the OPLVLT) while taking into account the potential errors that could have taken place (Rudner & Schafer, 2001).

Simply put, "the reliability of a measure refers to the consistency with which it measures whatever it measures" (Foxcroft & Roodt, 2001: 41). The reliability of the OPLVL Test was established by calculating the Cronbach coefficient alpha, Pearson r and Spearman-Brown formula which provides the ratio of a true score variance to the observed score variance, calculated as $r= \sigma T/\sigma X$ (Foxcroft & Roodt, 2001; Gregory, 2007).

5.1.1. Internal Consistency

Internal consistency was identified by Cronbach (1951) to be the prominent measure of reliability. Internal consistency focuses on the degree to which the individual items are correlated with each other and is thus often called homogeneity (Rudner & Schafer, 2001). This method of measuring reliability does not directly give an estimation of reliability but the theoretical estimate derived from classical test theory (Henson, 2001). Internal Consistency relates to item homogeneity or degree to which the items on a test jointly measure the same construct, the item interrelationship.

In terms of these assumptions, the items in the OPLVLT test must be highly interrelated because they assess the same construct of interest (vocabulary) meaning the items in the test should correlate highly with each other if they truly represent appropriate content sampling therefore if items are highly correlated, it is theoretically assumed that the construct of interest has been measured to some degree of consistency (i.e. the scores are reliable) (Anastasi, 1997).

The Internal Consistency of a scale should ideally be 0. 95 (Nunnally, 1967), or at least 0.80 according to Loo (2001) it was estimated by reporting Cronbach's alpha. The Cronbach alpha was reported seeing that it is considered applicable to multiple-scored items (Anastasi, 1997). The OPLVL Test consisted of 36 items. A total of 120 learners responded to all the items in the test. Since this is a form of an achievement test (vocabulary achievement) and certain decisions are made from this score (see discussion in Chapter 6) the reliability coefficient should ideally be 0.8 and above (range is between 0.00 and +1.00) (Rudner & Schafer, 2001). The Internal consistency reliability of the test indicated by the Cronbach Coefficient Alpha was 0.84 indicating a high coefficient, indicating that the OPLVLT scale is a reliable instrument capable of producing consistent results.

Table 5.2 Cronbach's alpha for Internal reliability

Cronbach Coefficient	
Raw Scores	.84
Standardized scores	.86

The Cronbach coefficient was relatively high at .8688 as theories indicated above had suggested, and appeared to suggest strongly that the OPLVLT is a reliable instrument.

The Internal consistency of the Metropolitan on this sample yielded a Cronbach alpha of .80 the Peabody had an alpha of .77 both which was adequately high.

5.1.2. Item Analysis

Item Analysis was performed on the OPLVLT in order to ascertain the usefulness of each of the items on the scale, so as to inform decisions about Items that should be retained or dropped on the scale. As the table below indicates, most of the items on the scale had positive correlations with the Total score of the test and the OPLVLT had an overall Internal reliability of .84 which indicated an estimated 16% measurement error in the OPLVL Test.

Further reliability analyses were conducted, they involved identifying items with low correlations (>0.5) with the entire instrument and calculating whether removing these would significantly change the reliability coefficient of the scale. Item 1 which produced a correlation of -.04 was the example item when administering the test and all the learners scored correctly on it, hence the low correlation was removed and surprisingly had no significant effect on the overall reliability of the instrument which remained high at .8679.

Item 2 also had a low correlation with the overall instrument (i.e. .03) and when removed the overall reliability dropped slightly from .8688 to .8662. Item 20 had a low correlation of .08 and dropped the overall reliability slightly to .8651. Item 34 also only slightly dropped the reliability from .8688 to .8681 although it had a low, negative correlation of -.05. The remaining 32 Items on the instrument had alpha's ranging between 0.83 and 0.85 when deleted and thus they were all retained for the test (see Table 5.3. below).

The item-total correlations for items on the OPLVL Test with the total of the scale ranged between .04-.48 (94.27%). According to De Vaus (2002), this would indicate that the items in the scale discriminated better for learners who attained higher scores on the overall test than it did for those who generally scored low on the test (De Vaus, 2002). Only 2.77% of the Items had correlations that were above 0.5 with the total score. This indicated that learners who generally did well on the test (i.e. learners who had a good vocabulary) scored correctly on this item (See Table 5.3 below).

Table 5.3: Inter-Item correlations and corresponding alpha's

	(Raw) Correlation	**Alpha**	**(Standardised) Correlation**	**Alpha**
Item 1	-.05	0.84	-.04	0.86
Item 2	0.04	0.84	0.03	0.86
Item 3	0.16	0.84	0.17	0.86
Item 4	0.23	0.84	0.25	0.86
Item 5	0.35	0.84	0.37	0.85
Item 6	0.48	0.83	0.50	0.85
Item 7	0.64	0.83	0.64	0.85
Item 8	0.20	0.84	0.22	0.85
Item 9	0.34	0.84	0.35	0.86
Item 10	0.58	0.84	0.58	0.85
Item 11	0.21	0.83	0.22	0.85
Item 12	0.17	0.84	0.14	0.86
Item 13	0.16	0.84	0.15	0.86
Item 14	0.53	0.84	0.54	0.86
Item 15	0.22	0.84	0.23	0.85
Item 16	0.40	0.84	0.39	0.85
Item 17	0.47	0.83	0.47	0.86
Item 18	0.23	0.84	0.23	0.85
Item 19	0.37	0.84	0.39	0.85
Item 20	0.09	0.84	0.08	0.85
Item 21	0.68	0.83	0.69	0.85
Item 22	0.59	0.83	0.60	0.85
Item 23	0.16	0.85	0.19	0.86
Item 24	0.20	0.84	0.18	0.86
Item 25	0.42	0.84	0.42	0.85
Item 26	0.42	0.84	0.44	0.85
Item 27	0.73	0.83	0.73	0.85
Item 28	0.57	0.83	0.57	0.85
Item 29	0.57	0.83	0.57	0.85
Item 30	0.61	0.83	0.63	0.85
Item 31	0.59	0.83	0.38	0.85
Item 32	0.38	0.84	0.38	0.85
Item 33	0.26	0.84	0.25	0.86
Item 34	-.06	0.85	-.05	0.86
Item 35	0.36	0.84	0.34	0.85
Item 36	0.37	0.84	0.35	0.85

5.1.3. Item Discrimination

Item Discrimination is the index correlation of an item with the item total correlation which identifies the ability of each item to discriminate well-performing learners from poorly performing learners (Anastasi, 1997). This is thus the correlation between the right/wrong answers that learners get on a given item and the total scores that learners get on the overall test, values range between -1 and 1.

A *large positive correlation* means that learners with a rich vocabulary will get the item right and learners with a limited vocabulary will get the item wrong (Anastasi, 1997). This is a good indicator that the test is a good measure for vocabulary in that it reflects the true knowledge of vocabulary for that specific learner.

A low positive correlation implies that learners who get the item correct tend to do poorly on the overall test and good learners on test overall got the item wrong. A *negative correlation* indicates items with poor discriminatory value (Anastasi, 1997).

In the analysis of the OPLVLT, items 1 and 34 were found to have negative correlations with the Total score of the test *(-.05)* and *(-. 06)* (respectively) which indicates that they have poor item discriminatory value and are not useful for the test. Item 1, however, is an example item and was used to familiarise the test-takers with the test and so every learner was shown the correct option on this item.

It would thus be expected that this item would yield a low discriminatory value. However, Item 34 was not an example Item and remains problematic with an Item-total correlation of -.06. The negative correlation indicates a need to remove the item from the scale.

It should be noted, however, that the standardised Alpha of the overall scale is not compromised by the item's negative Item-total correlation. Thus, even though it has poor discriminatory value and thus does not assist the researcher in making decisions and conclusions about the learners who either scored correctly or incorrectly on it, Item 34 has no negative effect on the overall scale.

For this reason the researcher does not recommend that this item be dropped as it does not affect the overall reliability of the OPLVL test. The decision is thus to retain it, and keep the format of the test as is.

5.1.4. Item Difficulty

Item difficulty analysis provides the range of difficulty for tests (reference). Table 5.4 below provides the number of items in the OPLVL Test and their corresponding percentages.

Table 5.4 Difficulty range of Items of the OPLVL Test

Difficulty level	Range	Number of items	%
Very easy	.90 to 1.0	0	0
Easy	.70 to .89	1	2.77
Medium	.50 to .69	10	27.77
Difficult	.30 to .49	11	30.5
Very difficult	Less than 30	13	36

The Item Difficulty analyses provided the range of difficulty for tests and the number of items in the OPLVL Test and their corresponding percentages. Ideally, a test has to have a balance between 'easy', 'medium' and 'difficult' items (Foxcroft & Rood (2001). The OPLVL Test as demonstrated in Table 5.4 has an imbalance of 'difficult' to 'very difficult' items. Very few Items are of 'medium' difficulty (27.77%) on the test according to the range, and still fewer Items are 'easy' to 'very easy' (2.77%). More than 66% of the Items range from 'difficult' to 'very difficult' making the test have enough items of sufficient difficulty to adequately test the particular sample of learners tested.

All items in the OPLVL Test yielded alpha of above .83 and the overall Reliability of the test was .84 Raw and ***.86*** for the standardised alpha. It was thus decided that all Items will be retained for the scale as none of the items was affecting the reliability of the total scale. Item 1 produced a negative score, which was accounted for by the fact that Item 1 was an example question and all the children were given the correct answer to this question.

Item analyses for the Metropolitan test produced a standardised score of ***.80***, and none of the Items changed alpha to below ***.79***.

The Item Analysis conducted on the Peabody yielded a standardised Cronbach alpha of ***.78*** and none of the Items on the scale dropped down to below ***.75***.

The above analyses indicate that, for the standardisation sample tested, the OPLVLT was a highly reliable test. For this sample, the other tests used for construct validation also had adequate levels of reliability (reference). This implied that each of the three tests used for construct validation purposes in this study were measuring single as opposed to multiple constructs.

The high levels of reliability in each of the three instruments also indicated that it would be possible to use all three instruments in an analysis designed to establish whether the construct measured by each test was discrete as an entity, or whether it overlapped with the constructs measured by the other tests.

The analyses conducted construct validation purposes are reported in the following section.

5.2. Validity

5.2.1. Construct Validity

A construct is some postulated attribute of people, assumed to be reflected in test performance. In test validation the attribute about which we make statements in interpreting a test is a construct. This validation takes place when an instrument (i.e. OPLVLT) is suspected to reflect a particular construct (i.e. vocabulary). A score produced by the OPLVLT is expected to indicate a possession or lack of the possession of the quantitative attribute (i.e. vocabulary) (Coulacoglou, 2002). According to Anastasi & Urbina (1997), correlations between a new test and an earlier test that has high construct validity and measures the same construct can be evidence that the new test measures approximately the same construct, hence the use of the PPVT and MRT School Language and Listening subtest.

Specific testable hypotheses were made; the first was how well the OPLVLT measures the same construct as the Metropolitan Readiness School Language and Listening Subtest which has high construct validity and the second hypothesis is how well the OPLVLT measures the same construct as the PPVT which also has high construct validity for vocabulary. Results of testing these hypotheses will confirm or disconfirm the claim that OPLVLT measures vocabulary by indicating as definitely as possible the degree of validity the test has on the sample it was administered to (Cronbach, 1955).

The correlations between the OPLVLT and the Metropolitan School Readiness School Language and Listening test were as follows;

Pearson Correlation 'r', N=120

Table 5.5: Summary of scores on the validation tests

	Score on Metropolitan	**Score on Peabody**	**Score on OPLVL**
Score on Metropolitan	1.00	0.45	0.29
Score on Peabody	0.45	1.00	**0.17**
Score on OPLVL	**0.29**	0.17	1.00

The analysis yielded a correlation coefficient of **r=.29** between the OPLVL Test with the Metropolitan School Readiness Vocabulary subtest and a correlation of **r=.17** with the Peabody Picture and Language Test (as shown in Table 5.5 above), these positive, yet small correlations indicate a low level of correlation between the tests. This indicates that although the tests are similar at face validity level in that they all state that they measure vocabulary and language, they appear to be measuring constructs which have little relationship with one another. Alternatively, that other confounding variables influenced the test scores.

The Cronbach Coefficient Alpha for the Construct Validity of the OPLVLT yielded the following;

Table 5.6: Cronbach Coefficient Alpha

Raw Scores	0.54
Standardized Scores	**0.57**

The overall Cronbach Coefficient Alpha between the three scales is a positive, moderate .54 (raw) and .57 (standardised) (see Table 5.6 above). Thus, the OLPVL test has weak associations with a previously standardised, reliable, valid and normed vocabulary test (the Peabody) and with a previously standardised, reliable, valid and normed school readiness language test (i.e. the School Language and Listening subtest of the Metropolitan Readiness Test and the Peabody Picture-Language Test).

Overall, the analyses suggest that the OPLVLT has a moderate level of Construct Validity in relation to the other tests used as part of the test battery (Cronbach's alpha of 0. 57).

It can thus be concluded that the OPLVLT appears to be school readiness related, but to measure a specific vocabulary construct not measured by the other tests.

5.3. ANOVA

5.3.1. Parametric ANOVA

One-way Analysis of Variance between two groups measures the difference between the means of scores of two groups on a dependent variable (Iversen & Norpoth 1987). The analysis of variance with gender as an Independent variable and score on the OPLVLT yielded the following results;

Table 5.7: Descriptive Statistics for Gender as an IV

Gender	N	Mean	Std. dev	Lower Bound	Upper Bound	Minimum	Maximum
1 (Boys)	37	32. 75	2.80	31.82	33.69	25.00	36.00
2 (Girls)	82	33.24	3.26	32.52	33.96	15.00	36.00
Total	119						

The Descriptive table above indicates that for 119 learners, gender was indicated and these were valid cases included in the analysis. Of the 119, 37 learners were male and 82 were female. Their group means on the OPLVLT were 32.75 for boys and 33.24 for girls. Minimum scores were 25 for boys and 15 for girls on the OPLVLT and maximum scores were 36 for both groups.

Test for homogeneity of groups

A test calculating whether the variance within the populations from where the samples have been drawn is equal was calculated using Levene's test. At 0.871 the test indicated a non-significant p-value implying that the null hypothesis which stated that the groups have been drawn from populations with equal variance.

Table 5.8: ANOVA for Gender

Source	Df	Sum of Squares	Mean Square	F Value	Pr>F
Between Group	1	6.05	6.05	.61	.433
Within Group	117	1145.93	9.79		
Total		1157.98			

As indicated on the descriptive summary the group sample means between the two groups are different at 32. 75 for girls and 33. 24 for boys. In establishing whether this difference was attributable to chance or to differences within the populations from which these samples came the F-ratio was calculated.

F yielded a reasonably low 0.61 value at 5% with a p-value of significance of 0.433 (which is not significant at 5%) indicating that the differences between boys and girls is likely due to chance and that the two populations are similar with equal means. The null hypotheses could not be rejected and this is further supported by the non-significance p-value at 5% which indicates that the differences between the two populations are not statistically significant.

Further analysis of the group differences were conducted on the background that the data should be examined for normality, homogeneity and independent. Hence, the non-parametric test, conducting the same analysis was performed.

5.3.2. Non-Parametric ANOVA

The first assumption made in ANOVA is that the data is normal. Normality tests are used to determine whether data set is well-modeled by a normal distribution. The main types of tests for normality are; a) Empirical Distribution tests: Shapiro-Wilk, Lilliefors and Kolmogorov-Smirnov, b) Skewness and Kurtosis tests, c)Regression and correlation tests and d) Normal probability plot. In this analysis Shapiro-Wilk and Kolmogov-Smirnov tests and the Normal probability plot were used. These analyses yielded the following results (with degrees of freedom at 121);

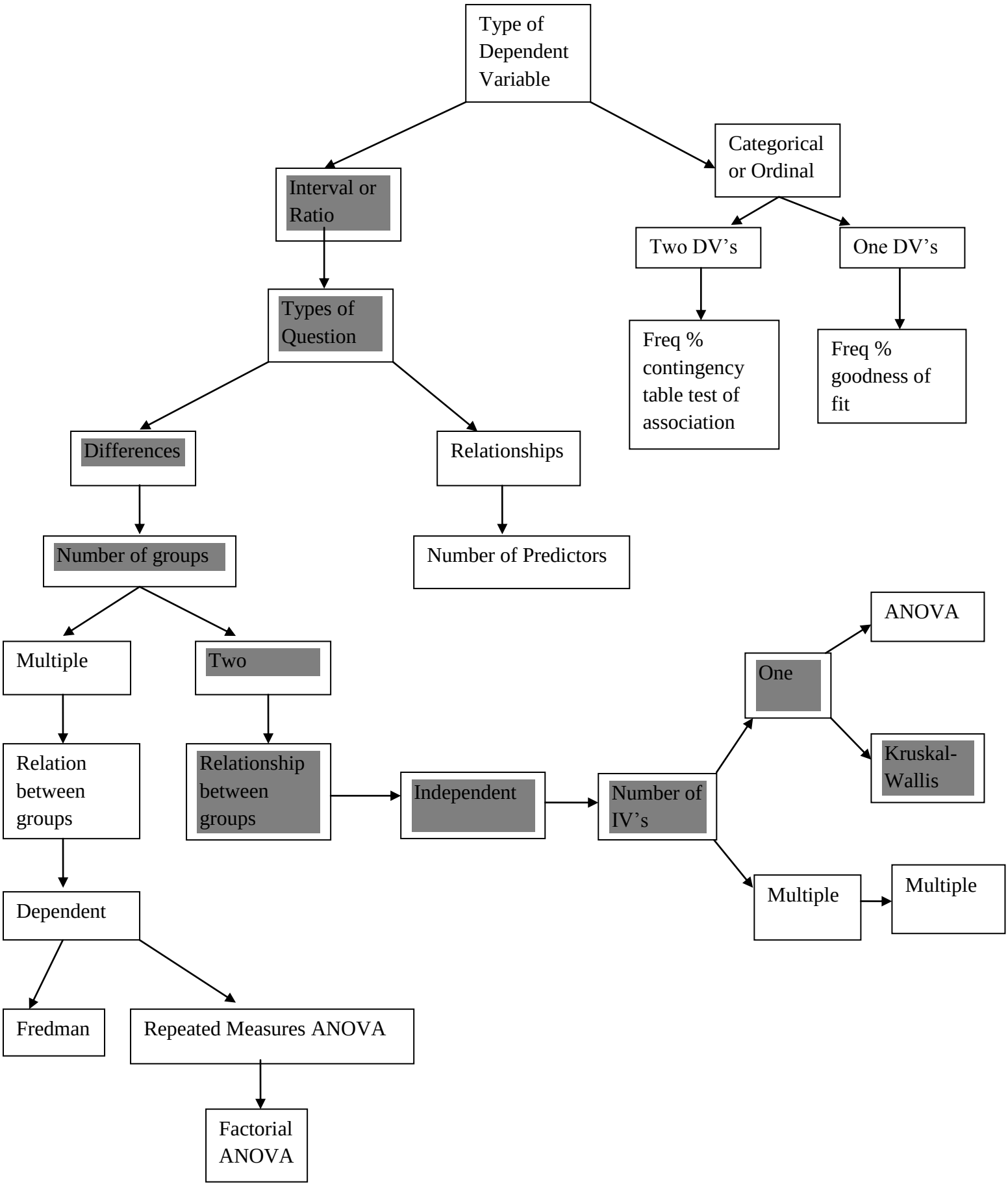
Table 5.9.: Non-parametric significance for Gender

	Kolmogorov-Smirnov		**Shapiro-Wilk**		
	Statistic	Sign		Statistic	Sign
Score on Metropolitan	.105	.062		.980	.067
Score on Peabody	.061	.200		.986	.261
Score on OPLVLT	.220	.000*		.729	.000*

The "null hypothesis" is that the data is normal and the "alternative hypothesis" is that the data is non-normal. Using the p value, indicated in the significance column, the p-value for the OPLVLT test was $p = 0.000$ which is less than alpha (level of significance) of 0.05. The null hypothesis that the data is normal was thus rejected, meaning that the OPLVLT scores were not normally distributed.

The probability plot transforms the data into a normal distribution and plots it as a scatter diagram. Normal data was expected to follow the trend line and non-normal data was expected to have more points farther the trend line. Please See Appendix H on page 103.

The statistical decision tree by Neill (2008) based on Howell (2008) employed in this study is illustrated below. The decisions pertaining to the methodology used were informed the Figure 5.1. The Statistical Decision Tree on the next page.


Type of Dependent Variable
Interval or Ratio
Categorical or Ordinal
Two DV's
One DV's
Freq % contingency table test of association
Freq % goodness of fit
Types of Question
Differences
Relationships
Number of groups
Number of Predictors
Multiple
Two
ANOVA
One
Kruskal-Wallis
Relation between groups
Relationship between groups
Independent
Number of IV's
Multiple
Multiple
Dependent
Fredman
Repeated Measures ANOVA
Factorial ANOVA

Test for homogeneity of groups

A test calculating whether the variance within the populations from where the samples have been drawn is equal was calculated using Levene's test. At 0.871 the test indicated a non-significant p-value implying that the null hypothesis which stated that the groups have been drawn from populations with equal variance.

Non-parametric Analysis of variance between two groups measures the difference between the means of scores of two groups on a dependent variable (Iversen & Norpoth 1987).

The analysis of variance with gender as an Independent variable and score on the OPLVLT yielded the following results;

Table 5.10: Non-Parametric Kruskal-Wallis ANOVA for Gender

Chi -square	Df	P-value
1.8819	1	.1701

As indicated on the descriptive summary the group sample means between the two groups are different at 32. 75 for girls and 33. 24 for boys. In establishing whether this difference was attributable to chance or to differences within the populations from which these samples came the chi-square was calculated.

Chi-square 1.8819 which was not significant at 0.005 as it yielded a statistically non-significant p-value of .1701 (which is not significant at 5%) indicating that the differences between boys and girls is likely due to chance and that the two populations are similar with equal means using both parametric and non-parametric measurements. The null hypotheses could not be rejected and this is further supported by the non-significance p-value at 5% which indicates that the differences between the two populations are not statistically significant.

Additional analyses of the influence of other variables on the scores were calculated employing parametric tests (ANCOVA) as these particular parametric tests can analyse more complex relationships and interactions between variables, and no difference has been shown between the results of both parametric and non-parametric tests conducted on the same results.

5.4. ANCOVA

The Analysis of covariance

The ANOVA results in section 5.3. indicated that gender had not been a significant influence on the results of the OPLVLT. In addition, the researcher investigated the effect of age on the scores as a covariate, with gender as a classifier and scores attained as a dependent variable.

Table 5.11: ANOVA for Gender and Age

Source	Df	Sum of Squares	Mean Square	F Value	Pr>F
Model	3	122.6158	40.8719	4.65	.0043
Error	107	940.8616	8.7931		
Corrected Total	110	1063.4774			

Table 5.11 above represents the combined effects of the two explanatory variables, namely; gender and age. The sum of squares is 122.6158 and the residual is 940.8616. The null hypothesis tested was that the sample means are the same with the population means. The F-value is high at 4.65 and is significant at the 5% level of significance with a p-value of .0043. The probability of getting this F-value when age and gender have no effect on the score is as little as .0043; as a result the null hypothesis of equal means is rejected.

Table 5.12 below has the effect sizes of the explanatory variables calculated separately, measuring the strength of each variable. These statistics indicate that Gender as a factor (F-value=1.41 and p=.23) is not significant and we fail to reject the null hypothesis stating that gender is the same in the population from which the sample was drawn from as in the sample itself. Age however, appears to be statistically significant at the 5% level of significance with an F-value of 4.56 and .03 p-value. The interaction between Age and Gender yields a non-significant effect on the OPLVLT score with a small F-value of 1.26 and a high probability of .2644 that the interaction of effects have no effect on the score.

Table 5.12: Strength of the effects on score

Source	Df	Type III SS	Mean Square	F Value	Pr>F
Gender	1	12.3939	12.3939	1.41	.2378
Age	1	40.0644	40.06442	4.56	*.0351
Age*Gender	1	11.07052	11.07052	1.26	.2644

R-Square: 0.1152, Coeff Variance: 8.9540, Root MSE: 2.9653

The Sum of Variance as indicated in Table 5.12 indicates that 12.39 of the variance in the OPLVLT test score is the effect of gender, this effect is not significant at 5% although it has a reasonable F-value. Age, however, is significant (p=.03) explaining a variance of 40.06 and the interaction between age and gender combined is not significant at 5% with a variance of 11. 07. This result informed the decision to look at age as an effect on the score without gender as it proved to be more influential than gender or than itself and gender interacting. The results yielded from this analysis are as follows;

Table 5.13: ANCOVA for Interaction of age and gender.

Source	Df	Sum of Squares	Mean Square	F Value	Pr>F
Model	2	111.5452	55.7726	6.33	.0025
Error	108	951.9321	8.8141		
Corrected Total	110	10634774.			

R-Square: 0.1048, Coeff Var: 8.9647, Root MSE: 2.9688

A separate analysis supported the results indicated in Table 5.13 of a significant effect of Age on the OPLVLT score. Age yielded a high F-value (11.96) and a significant p-value of .0008.

Table 5.14: Non-parametric ANOVA for Age and Gender as separate variables if the data were normally distributed;

Source	Df	Sum of Squares	Mean Square	F Value	Pr>F
Gender	1	4.8343	4.8343	0.55	.4605
Age	1	105.4366	105.4366	11.96	*.0008

Table 5.15: Parameters for Age and Gender

Parameter	Estimate	Standard Error	t-value	Pr>|t|
Intercept	42.1382	2.5847	16.30	<.0001
Gender 1	-.4493	0.6066	-.74	.4605
Gender 2	0.0000	.	.	.
Age	-1.2616	0.3647	-3.46	.0008

Table 5.15 above indicates the predictions for scores for girls and boys separately and for the different ages. ANCOVA has a regression component which allows us to make predictions and they are calculated when we have all values of y and need to estimate the score of a particular individual (boy or girl) of whatever age.

Boys score lower when they are older as the negative relationship of their estimate indicates (i.e. -.4493). Girls however, improve their score with age, indicated by the positive estimate. Age is therefore has a significant effect (p=.0008) on score differently for boys as it does for girls. The regression line for boys will consequently be;

Boy: Score=41.69-1.26*Age

Girl: Score=42.12=1.26*Age

These regression lines will allow for the prediction of the score of either gender by featuring in their age.

An Analysis of variance for the Research Question 4 was to be conducted to test the hypothesis of differences in scores due to first language. However, as a result of the systematic randomised sampling method used there were significantly fewer English first language speakers in the sample (i.e. five). This resulted in the inability to conduct any statistically appropriate comparisons. This Research Question was thus left unanswered.

5.5. Summary of results

This chapter aimed to answer the research questions set out in this study. Research question one asked if the OPLVLT test is reliable and section 5.1 provided evidence of the reliability of the test through yielding a coefficient of 0.86. Following Cronbach (1988) and Anastasi (include up to date reference) this level is more than adequate.

The second research question asked whether the OPLVLT is a valid testing instrument. The levels of first order correlation between the three tests (the OPVLVLT, the Peabody and the Metropolitan) indicated that although the tests appeared to be measuring constructs which have little relationship with one another. Nevertheless, at 0.57 Cronbach's alpha for Construct validity indicated a low overall relationship between all three tests (the OPLVLT, the Peabody and the Metropolitan).

This result indicates one of two things; that the OPLVLT is not a valid instrument for vocabulary or that the OPLVLT is measuring a different aspect of vocabulary, different from

the one measured by the Metropolitan and Peabody. Alternatively, another explanation could be that other confounding variables influenced the test scores. These and other possible interpretations will be discussed in Chapter Six.

The third research question investigated the relationship between gender and the scores attained on the OPLVLT and whether these two variables are related. The results indicated that although boys and girls scored differently on the tests, the differences between the two groups were not statistically significant, indicating that these differences were not due to differences between their populations. The differences were thus likely to have occurred by chance for this particular sample.

Research question four was left unanswered due to the small number of learners who indicated that they spoke English language. The nature of the data made it statistically inappropriate to conduct any comparative analysis using this group considering the large groups of the languages.

The final question, querying whether age may also have influenced the scores attained on the OPLVLT test indicates that age does have a significant effect on scores. This affects boys differently to girls. The effect is negative for boys, that is, the older the boys are the less their scores tend to be for girls however, the relationship of the effect on the score is positive which translates to meaning that the older the girls are the better they perform. The interaction of age and gender was not significant. When gender as a variable was analysed separately, this had no effect on the scores attained by the learners.

Chapter Six discusses the implications of these results on decisions concerning the use of the OPLVLT test as it stands, the conclusions about its reliability and validity and recommendations of improving the instrument.

CHAPTER 6

DISCUSSION

Messick (1989), Mislevy et al., (2002) and Kane (1992, 2001), understand scores on language tests not simply as reports of performance in the test situation but as having meaning only as the basis for inferences beyond that situation. The real target of the assessment, the candidate's performance under non-test conditions, cannot be observed directly but is inferred from observations of test performance. This chapter discusses the factors that influenced the test scores and consequently, the validity of the test; the inferences we can make about the samples vocabulary; the implications of the test as it stands; the conclusions about its reliability and validity and recommendations of improving the instrument (i.e. OPLVLT) and other recommendations for test developers.

6.1. Extraneous variables that may have influence the test scores;

6.1.1. The formats of the tests

While low correlations were noted between the OPLVLT and the two criterion tests (the Peabody Picture-Vocabulary Test and the Metropolitan Readiness Cronbach Coefficient Alpha for the Construct Validity of the OPLVLT when correlated with two other tests measuring vocabulary yielded a positive, moderate value of 0.57. Thus, though the OPLVLT did the OPLVL test has adequately strong associations with reliable, valid and normed vocabulary tests (i.e. the School Language and Listening subtest of the Metropolitan Readiness Test and the Peabody Picture-Language Test), confirming that is does indeed measure vocabulary as it is sufficiently correlated to other tests measuring the same construct 'vocabulary'.

This is due in part on the format of the tests and the varying ways of measuring the constructs, at face value (face validity) the tests look very identical in that they are Multiple Choice Format (also known as MCQ's) and they were all printed on Black and White photocopied paper ensuring that none of the pictures were more clearer than the rest. However at closer introspection there a number of differences that could have contributed to the small correlations.

The first explanation might be that although all three tests used MCQ's, they were structured differently. The PPVT and OPLVLT had four options given for each question whereas for the

MRT some of the questions (12 of the 16 questions) had 3 options and others (4 of the 16 questions) had four options.

Secondly, the structure of the questions was also different between the PPVT and the OPLVLT as opposed to the MRT. Whereas the PPVT and OPLVLT both had the first questions with *'mark picture of a tractor'* and repeated again, *'mark the picture of tractor'* this instruction is shorter when the instructor has said it many times because the test-taker then understands that they will have to mark the picture and they only really focus on what picture they have to make the mark. Such that the instruction is something simple like *'tractor'* by the 5th question. MRT's instructions are slightly different; they are longer, and are in a form of a narration. An example is '*Put your finger on the HAT. Listen. Jill lost a button from her coat while she was playing. Her mother took the coat to fix it. She got out a needle, some thread, and a pair of scissors. Mark under the picture that shows what Jill's mother still needed'*. The narration format demands a number of skills; such as listening, attention, concentration and the skill to follow the story. If the learner's were distracted about anything they may miss a key word and select an incorrect option. This test is thus far more demanding than the OPLVLT and PPVT whose instructions are simply 'boy' after the learners have understood that they need to mark under the picture that corresponds to the stimulus word given by the instructor. This difference may account for the small correlations between the three tests.

The third possible explanation for the small correlations and varying scores are other instructions on the tests. The MRT Test has two added instructions as compared to the PPVT and OPLVLT. In the MRT test-takers are given specific and different instructions for each question and each page. Each page of the test has pictures at the top of the page and the test-taker is first instructed to open to that page, and then instructed to find the correct row of question there that they need to look at. An example of this instruction is '*Turn your booklet to the next page. Then fold it so that this page is on top. Here you see some RED STARS at the top of the page. Put your finger on the ball. Listen.'* This may add to the cognitive demands of the test-taker such that they may begin to experience some anxiety when they even as much as fail to get to the page on time or struggle with identifying the pictures corresponding to the question. However, due to the fact that the aim of this analysis was no to indicate that the MRT is correlated to the OPLVLT but to indicate that the OPLVLT is as reliable as the MRT, this was well adequately demonstrated.

The other factor that could possibly be contributing to this is the size of the pictures. The pictures in the three tests are of differing sizes. In the PPVT test for example each A4 size page has four pictures with each picture the same size as the others on that page. This test has the biggest pictures. The OPLVLT has smaller pictures than the PPVT and fits 24 onto one A4 size page, the MRT fits at most, 16 pages onto an A4 size page, but because the pictures have more than one object (they have up to 7 objects on one picture) this makes the pictures smaller. This difference in the size of the picture may disadvantage those learners who have poor vision and may require longer time to identify the stimulus object. The boundaries between the pictures in the MRT were very unclear and it was observed that a number of times the learners did not make out where one picture began and where another ended. This too, could have possibly led to varying scores on the tests even though the tests measure the same construct.

6.1.2. Socio-Economic Status

The Socio-Economic Status (SES) of the learner's family and the school environment in general such as the availability of learning material and resources play a significant role in vocabulary learning. Some findings suggest that low-SES English Language learners with limited vocabulary should be provided with early assistance in vocabulary acquisition. Gottardo et al. (2008) suggests that these interventions should take place prior to the time when these skills are crucial for reading comprehension because vocabulary scores do not seem to improve spontaneously. This essentially means that these interventions need to be provided quite early on as early as the first three grades of school as other studies have indicated good vocabulary knowledge might independently enhance greater ability to learn new vocabulary and thus improve reading ability as has been argued elsewhere (Geva & Farnia, 2005).

The learners who participated in this study came from schools which appeared to have had equal access to resources. The classrooms were covered in pictures and there appeared to have been enough reading material in each of the schools. What the researcher could however not elicit was whether the learners had the same access to these resources at their respective homes. Theories of vocabulary have indicated that exposure to rich linguistic input enhances the quality and speed of learning language and perhaps also contributing to the performance of the learners was their level of exposure to language input outside the school.

6.1.3. Test Anxiety

Test anxiety measures anxiety reactions in academic testing situations and appears to be inversely related to test performance (Mandler and Sarason 1952); thus, the more anxious a person is, the lower their test performance tends to be. Hill and Wigfield (1984) claimed that almost 25% of the testing population has some form of debilitating test anxiety. Other similar psychological matters that affect performance are Inattention and Non-response; inattention refers to being easily distracted and eventually selecting options carelessly (Wright, 1977) and non-response refers to not choosing any or all options available. A report from the test administrators in this research indicates that a number of learners selected more than one option when they were unsure of the correct option. Other learner's would look to the administrators to get some confirmation about which items to choose and would then select a number of options. Upon going through the test booklets, 16 learners had done this in at least one question, consequently these items were marked incorrectly. During the test taking process some factors may have contributed to raising or alleviating anxiety. These are; (i) the environment, (ii) the test administrators, and (iii) how much the learner thinks they know about the subject.

(i) The Environment

The learners were tested in their familiar environment, their schools, and for some of them, they were tested in their classrooms and therefore they did not have to deal with any unfamiliar environment. Because each school included in this study had a number of Grade 1 classes, the learners were selected randomly from each class and mixed with their fellow classmates in one Grade 1 classroom, making some of the learners remain in their classrooms and others move. The classrooms remained familiar to them in that they had been in them before.

(ii) Test administrators

The test administrators were trained, third year psychology students. They had acquired adequate test taking skills from the University of the Witwatersrand in their Research, Design and Analysis course. The administrators were therefore younger than the average teacher figure that the learners were used to, ranging between 20 and 25 years of age. The researcher identified this age factor as may have made the situation less threatening for the learners. There were also icebreakers performed before each test taking session to relax the learners and this was may have also contributed to alleviate the anxiety levels of the learners. Perhaps,

however, the strangeness of the administrators contributed to some level of anxiety as they were not people that the learners had seen before. In addition to the unfamiliarity to the test administrators, the sitting arrangement in the classroom may have possibly been an anxiety provoking element, because the learners normally sitting in groups of four to six, facing each other. During the test administration, the learners were sitting individually, all facing the front and not being allowed to talk to anyone in the classroom.

(iii) The learners' prior knowledge

Learner's who are not confident in their possession of the attribute being measured were expected to have relatively higher levels of anxiety than their counterparts who do. This hypothesis was, however, not tested.

6.1.4. Studies of group differences among test-takers

The test takers were different on a number of variables such as age, race, gender, school and first language and similar in that they were all in the first grade of school for the first time. These differences are indicated in the validation sample section in the Chapter Four. They are summarised in the table below;

Thirty seven of them were boys (30.83%) and Eighty-three of them were girls (69.17%). They came from three schools around Johannesburg; Welizibuko Primary School (uses English as the language of learning and teaching) in Kliptown, Soweto (n=43); Paul Mosaka Primary School (uses seSotho as the language of learning and teaching) in Pimville, Soweto (n=43); and Bree Primary School (English language of learning and teaching) in Mayfair (n=43) Primary Schools. The learners mean age was 7.02 years old and their age ranged between 6.0 and 11.0 years of age. The learners spoke various home languages, with 37 (30%) of the 120 learners having indicated that they speak seSotho as a first language, 35 (29. 1%) of the learners speak isiXhosa as a first language, 12. 50% (15) speak isiZulu as a first language, 11.66% (14) spoke Setswana and 4.16% of the learners reported to be speaking English as a first language.

These differences also somewhat explain the observed differences in the learners with regard to their performance on the test. The four factors described above; group differences, socio-economic status, the test differences and anxiety attempt to provide explanations for the variations on the test scores.

6.2. Implications of the test as it stands

6.2.1. The construct measured by the test

Definitions of vocabulary or knowing a word were discussed in the literature review in Chapter Three, the definitions presented varied widely, however, different theorists have reached consensus about these two principles; these were that vocabulary should include size and quality and not just possessing information about what it means to know a word but also knowing the context in which it can be used and having it accessible for use whenever needed. However, the nature of the vocabulary tests administered in this research do not measure the quality of the knowledge of a word, but only the ability to recognise the picture corresponding to that particular stimulus word, meaning, the size of vocabulary only. This was achieved through requesting learners to identify pictures corresponding to the stimulus words presented by the test administrator. This only measures a superficial understanding of the word which is expected at the level of the test-takers who participated in this study and thus was age and grade appropriate. The test scores did not indicate the amount of vocabulary that the learners knew, but were indicative of the discrimination between the sizes of vocabulary the learners possess. The understanding of a word was thus measured only by the mastery of matching the picture to the stimulus word presented. Therefore, this test can be said to be measuring this aspect of vocabulary.

6.2.3. Discrimination power of the OPLVLT

The strength of any achievement test lays in its ability to discriminate the scores of those who possess the attribute being measured and those who do not, hence, the use of standardised tests. Standardised tests provide the test administrators with a normative score to compare each test-taker on (http://en.wikipedia.org). The sample used for standardising the OPLVLT assisted in creating the normal distribution which can be used for comparisons of any specific future scores (also known as norm group). In validating the OPLVLT there was inevitably a standardisation of the OPLVLT; as mentioned earlier however, this study constitutes the piloting of the standardisation, not the national standardisation, due to the small sample size. However, regardless of the small sample size the test scores could be standardised for the learners who participated in the study and perhaps those with similar characteristics including their classmates who did not necessarily participate. The scores attained in the study were thus normed on the characteristics of the learners who participated such as their grade, age, language used as language of learning and teaching and home language.

Table 6.1. Normed scores on the sample

Variable	Raw Score Mean	% Mean	Std dev	Min	Max
Metropolitan	8.63	53.93	3.65	0	16
Peabody	21.90	62.57	5.67	9	35
OPLVL	33.08	91.88	3.13	15	36

The average scores on the different tests as illustrated on Table 6.1. above could then be used as the benchmark for rating each learner. The PPVT and MRT had been previously shown to have high discriminatory properties in terms of distinguishing between learners who have high vocabularies and language skills with those who do not (confer with Dunn & Dunn, 1997 and Swanson, 1981). The OPLVLT proved to be very sensitive in discriminating between learners who attained higher scores on the overall test than it did for those who generally scored low on the test.

Overall, the test has good, sufficiently strong validity and taking into cognisance the variability of the construct vocabulary as measured by the validating instruments.

6.3. Limitations

The most prominent limitation of this study came about as a result of the stratified sampling method employed by the researcher. This method involved selecting every third name of a learner from their class registers into the research sample, as indicated elsewhere; this method resulted in having unbalanced numbers between the two genders and the language groups. Gottardo et al. (2008) has identified the weakest link in the reading abilities of second language children to be vocabulary; this is the lack of knowledge of the words they are reading. For English Speakers (ES') this is not so much an issue as it is for English Learners (EL's) because of their exposure to the English language beyond the classroom. At one of the schools where the data for this study was collected, Bree Primary School in Mayfair, there were 5 English Language Speakers, who have evidently faired far above their fellow EL's both from their school and the other two schools. English first language speakers are said to have wider breadth and depth of vocabulary than English second language learners (Gottardo et al., 2008; Geva & Farnia, 2005) which essentially means that the vocabulary size of English language speakers (ES's) is typically much larger than that of second language speakers (EL's) (Kunnan, 1998). The ES' at Bree Primary scored an average of 35.2 on the

OPLVLT, 29.6 on the PPLT and 14.00 on the MRT, whereas their fellow classmates scored relatively lower, providing evidence for the above statement. The table below summarises these scores;

Table 6.2. Scores on tests by Language

Test	English	Urdu EL's	isiXhosa EL's	isiZulu EL's	seSotho EL's
MRT	14	9.57	6.22	10.73	8.9
PPVT	29.6	22	19.17	24.53	21.22
OPLVLT	35.2	34.7	33.4	32.8	32.66

It would have benefited the study's results to have been able to measure the effect of these different languages on scores statistically. However, due to the fact that only 5 of the learners who participated spoke English as a language this examination of effect was rendered statistically unfeasible. As a result, having only five learners who spoke English as a home language made statistically inappropriate to conduct any comparisons (i.e. ANOVA or t-test) using home language as a variable. This would have provided this study with considerably more explanation of vocabulary knowledge among first grade learners in the schools sampled.

No factor analysis was conducted in this research, which meant that no pairings of the independent variables were made and that each of the independent variables were looked at individually. This analysis could benefit the larger validation study to be conducted in the future.

A third potential limitation in this study is derived from Messick's (1989) framework which indicates that validation should not be once off but should be carried out over time; this is to ensure that changes that take place over time in terms of how many words learners know at a specific grade are captured. It would have been useful in this study if there was a time lag and a second testing in order to strengthen the validation of the test, this however, would have posed a number of other factors such as bias (the learners would have been exposed to the tests before, and would naturally perform better), and using a different sample would have not completely reflected the validity of the test as the scores would be different depending on the composition of the sample.

These limitations serve as recommendations for the validating study that will be undertaken as a result of this pilot study.

6.4. Recommendations

6.4.1. Recommendation for Vocabulary teaching

The learning process of vocabulary has a number of stake holders; the educators, the learners, the parents of the learners and the Department of Education to mention but a few. Each of these stakeholders has a tremendous role that they need to play in the successful learning process of the English vocabulary. Theories make several recommendations of ways that are believed to have the ability to enhance the process of vocabulary learning and I make mention of some of these below. Central to these recommendations is the need for a collaborative effort between all stakeholders and the acknowledgement and emphasis that vocabulary learning takes place in different aspects such as listening, reading, teaching, speaking and writing. All these elements must be integrated into the learning process. It appears at face value that learning words in another language is an easy, almost effortless process however; consistently research is proving that it isn't. Learners have the bulk of the work, beyond the instruction that is undertaken by the educator in the classroom, learners must create opportunities to apply their language skills in real life situations (Myburgh, No Year).

6.4.2. Factor of Language

Gottardo et al. (2008) has identified the weakest link in the reading abilities of second language children to be vocabulary; this is the lack of knowledge of the words they are reading. For English Speakers (ES') this is not so much an issue as it is for English Learners (EL's) because of their exposure to the English language beyond the classroom. However, some theorists have identified a number of possible solutions to this matter such as urging educators to encourage learners to assist each other (Myburgh, No Year) and assigning learners to the appropriate proficiency level group in an intensive English programme (Gottardo et al., 2008).

6.4.2. Recommendations for Test Developers

There are a number of lessons that could be learned from this pilot study. These include using a broad range of participants to highlight group differences, using young test administrators to alleviate anxiety of the test-takers and using various tests with various formats to measure the construct differently. Test developers may also benefit from using a larger sample to

ensure sufficient representativeness of each group in terms of group characteristics and combinations such as English first language male speakers. The method of random sampling; extrapolation, made it difficult for the researchers to ensure that each group is well presented hence the suggestion that the sample size be increased to increase the chances of having a broad and well representative sample. With regard to the analysis, it might be useful for test developers to consider using other statistical methods such as factor analysis in order to test the accuracy of the results.

A final recommendation is to administer more tests that measure the construct, such as the Kaufman Brief Intelligence Vocabulary subtest, the Verbal Intelligence Quotient (IQ) of the Wechsler Intelligence Scale for Children Revised Version, Receptive One-Word Vocabulary Test, Expressive Vocabulary Test and the Expressive One Word Vocabulary Test-Revised Version. The use of these tests can increase the validity of the test as it would require more than one faculty (that is, the identification of pictures from a stimulus word). The other tests listed require more than simply the identification of pictures, they include receptive and expressive vocabulary where a test-taker is assessed on their ability to listen or produce sounds in the language. This was not done in this study for a number of reasons including the amount of time it would require to administer all these tests and the assumed level of vocabulary size of the learners involved (most of whom had never been exposed to English before until they had started school four months prior to the testing). It could add to similar studies and the validation study to be carried as a result of this pilot study.

The Open Learning Systems Education Trust commissioned the study of validating the OPLVLT to form an essential component of their ‘English in Action’ programme by creating the vocabulary test to measure vocabulary in Grade One learners. Before this instrument could be widely used, it had to validated and this study successfully achieved the piloting of that process. The results of this pilot study indicate that the OPLVLT is a valid and reliable instrument to use and the results provide recommendations for the manual development of the instrument.

CHAPTER 7

CONCLUSION

Validation is about the quality of the interpretations made of test scores such as interpretations and inferences (Messick, 1989; Haladyna, 2004). Messick's validation framework adds that validation involves the accumulations of various kinds of evidence to support the desired interpretations of the test scores under construct validity (1989). This study examined the psychometric properties of the OLSET Picture-Language Vocabulary Level 1 Test as a pilot study for Grade One learners in three schools in Johannesburg to ensure the quality of the interpretations and inferences of the test. The learners used in the study varied on certain characteristics such as gender, age and home language and scores of the various groupings are reported.

The literature consulted confirmed a need for culturally appropriate vocabulary assessments instruments for Grade One learners in South Africa. This study aimed to bridge this gap by piloting the validation of a test developed in South Africa for use on samples with characteristics similar to this.

The process of validating the test involved calculating its reliability, validity and comparisons of the learners on gender and age. Reports of the reliability analyses indicated that the OPLVLT has 0.86 internal consistency reliability indicated by Cronbach's alpha. This reliability coefficient is relatively high and indicates high reliability properties for the scale.

The Validity analysis yielded an overall construct validity of 0.57 with relatively low correlations with the two other scales used in the validation calculations; these were 0.29 with the MRT and 0.17 with the PPVT. Reasons and suspicions for these correlations have been discussed in both the findings and discussions chapters of this paper. The overall Construct Validity of the test of 0.57 suggests that although the OPLVLT has overlap with other vocabulary and language tests to some degree but may not be measuring completely similar constructs of vocabulary to other commercially available tests. However, it is is doing so to some degree.

In terms of this pilot validation study conducted with this particular sample, further work with the OPLVLT with a bigger validation sample is indicated.

The central aim of this study was to establish the psychometric qualities of the new group test, OPLVLT, which was designed to be used to measure the level of vocabulary of Grade One learners. The instrument was shown to have adequate reliability coefficients, as well as a low overall construct validity with other tests/instruments measuring vocabulary and language, in terms of other vocabulary and listening-related functions. The test thus has sufficiently high construct validity to state that the instrument is indeed measuring vocabulary, though the facet of vocabulary it is measuring is apparently substantially different to the aspects of vocabulary and language measured by other tests.

More specifically, the study aimed to validate the OLSET Picture-Language Vocabulary Level 1 Test by providing evidence of its reliability and validity properties and conducting Item Analysis whose results would inform the retaining or discarding of some of the items. This was successfully achieved with the data collected at the sampled schools.

The second aim was to estimate the influence of possible third variables on the data, relative to the possibility of using the data set to establish standardised scores for the sample population. The reason for doing so was to demonstrate the effect of third variables such as gender, language and age on the scores attained by the sampled learners. Unfortunately, the language variable could not be measured due to the nature of the data, but gender and age were successfully estimated and reported.

All, except one of the Research questions were answered. The first Research question was whether the OPLV Level 1 Test a reliable measuring instrument which produces consistent and accurate scores. The answer was calculated and was that, yes; the OPLVLT is a reliable and consistent measuring instrument.

Research Question Two was whether the OPLV Level 1 Test measures vocabulary, and whether the Test a valid measure of the construct; vocabulary, as it claims to be? The answer gathered in this pilot study was that yes, the OPLVLT is a valid measure of vocabulary.

Research Question Three focused on the significance of the differences between the scores of male and female learners on the test. Differences were found to be present but not statistically significant.

Research Question Four related to differences due to language. This question could not be investigated and could therefore not be answered.

Research Question Five investigated differences due to age and this variable was found to have a statistically significant influence on the scores attained by the learners on the test.

Overall, this study indicated that the OPLVLT does not require any amendments and is a reliable instrument. Further validation is necessary for the development of a technical manual of the OPLVLT.

REFERENCES

American Psychological Association. (1985). *Standards for Educational and Psychological testing.* Washington, D.C: American Psychology Association.

Aiken, L. R., & Groth-Marnat, G. (2006). *Psychological Testing and Assessment, 12th Ed*.

Alderson, J. C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.

Anastasi, A. (1988). *Psychological Testing*. New York: Macmillan.

Anderson, R. C., & Freebody, P. (1981). "Vocabulary knowledge." In J. Guthrie (ed.), *Comprehension and Teaching: Research Reviews, 77: 117*.

Arnott, A., Mansfield, J. and Mentis, M. (1993) (a) 'Pre-Test Report of "English in Action"', in C.S Potter, and S. Leigh, (eds) *English in Action in South Africa 1992–1994*. Washington DC: Learntech and Johannesburg: Open Learning Systems Education Trust.

Arnott, A., Mansfield, J. and Mentis, M. (1993) (b) ' A Summative Evaluation of OLSET's "English in Action" Radio Learning Programme', in C.S Potter, and S. Leigh (eds) *English in Action in South Africa 1992–1994*. Washington DC: Learntech and Johannesburg: Open Learning Systems Education Trust.

Arnott, A., Mentis, M. and Mansfield, J. (1994) 'A Summative Report on OLSET's "English in Action" Radio Learning Programme', in C.S Potter, and S. Leigh, (eds) *English in Action in South Africa 1992–1994*. Washington DC: Learntech and Johannesburg: Open Learning Systems Education Trust.

Aukrust, V. G. (2007). Young Children Acquiring Second Language Vocabulary in Preschool Group-Time: Does Amount, Diversity, and Discourse Complexity of Teacher Talk Matter? *Journal of Research in Childhood Education, 22*, 17-37.

Barry, C. (2000). *The Applicability of the Peabody Picture Vocabulary Test-Revised, forms L and M, and The British Picture Vocabulary Test to 5-6 Year Old South African English Speaking Children*. Unpublished Thesis: University of the Witwatersrand , Johannesburg.

Blachowicz, C. L. Z.; Fisher, P. J. L. & Ogle, D. (2006). Vocabulary: Questions from the classroom. *Reading Research Quarterly, Vol. 41*, 524-539.

Blachowicz, C. L., Fisher, P. R. (2004). Vocabulary lessons. *Educational Leadership, 61*, 34-45.

Boers, F.; Murielle, D.; & Eychmans, J. (2004). Cultural variarions as a variable in comprehending and remembering figurative idioms. *European Journal of English Studies, 8*: 375-388.

Buck, G. (2001). *Assessing Listening*. UK: Cambridge University Press.

Bush, M. D. (2007). Facilitating the Integration of Culture and Vocabulary Learning: The Categorisation and use of pictures in the classroom. *Foreign Language Annals, 40*, 727-745.

Carey, S. (1978). Acquiring a single new word. *Child Language Development, 15*: 17-29.

Carver, R. L. (1997). Theoretical underpinnings of service learning. *Theory into practice, 36* (3)

Catts, H. W.; Bridges, M. S.; Little, T. D. (2008). Reading achievement growth in children with language impairments. *Journal of Speech, language and Hearing Research, 51*: 1569-1579.

Chang, A. C. (2007). The Impact of Vocabulary Preparation on L2 Listening Comprehension, Confidence and Strategy Use. *System, 35*, 534-550.

Chang, A.; Cheung, M. ; Leung, W. & Cheung, R. (2008). Measuring vocabulary by free expression and recognition tasks: Implications for assessing children, adolescents and young adults. *Journal of clinical and experimental Neuropsychology, 30*, 892-902.

Chang, A.; Cheung, M. ; Leung, W. & Cheung, R. (2008). Measuring vocabulary by free expression and recognition tasks: Implications for assessing children, adolescents and young adults. *Journal of clinical and experimental Neuropsychology, 30*, 892-902.

Chapelle, C. (1998). Construct definition and validity inquiry in SLA research. In L. F. Bachman & A. D. Cohen (Eds.), *Interfaces between second language acquisition and*

language testing research (pp. 32–70). Cambridge, England: Cambridge University Press.

Childers, J. B., & Tomasello, M. (2002). Two year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology, 38,* 967-978.

Ching-Shyang Chang, A. (2007). *The Impact of vocabulary preparation on L2 listening comprehension, confidence & strategy use. System, 35,* 534-550.

Crawford, L.W. (1993). *Language and Literacy learning in multicultural classrooms*. Massachusetts: Allyn & Bacon.

Croll, C. (1971). Teaching Vocabulary. *College Composition and Communication, 22*, 378-389.

Cronbach, L. J. (1942). An analysis of techniques for diagnostic vocabulary testing. *Journal of Educational Research, 36,* 206–217.

Cronbach, L. J. (1988). *Essentials of Psychological Testing, 2nd Ed*. New York: Harper & Collins.

Cronbach, L. J. and Meehl, P. E. (1955). Construct Validity in Psychological Test. *Psychological Bulletin, 52,* 281-302

Dale, E., O'Rourke, J. (1971). *Techniques of Teaching Vocabulary*. New Jersey: Field Educational Publication.

Davies, A. (2006). *Language Education: Standards Movement*. University of Edinburgh: United Kingdom: Elsevier Ltd.

Davies, A. (2006). *Language Education: Standards Movement*. University of Edinburgh: United Kingdom: Elsevier Ltd.

Dean, S. & Illowsky, B. (2008). F-Distribution and ANOVA: Purpose and Basic Assumption of ANOVA. *Connections, 1.5*.

De Vaus, D. A. (2002). *Analysing Social Science data*. London: Sage Publishers.

Deno, S. L. (1968). Effects of words and pictures are stimuli in learning language equivalents. *Journal of Educational Psychology, 59*: 202-206.

Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary-Revised Manual Forms L and M*. Minnesota: American Guidance Service.

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24*, 174-187.

Finocchiaro, M, (1969). *Teaching English as a Second Language*. New York: Harper & Row, Publishers.

Foxcroft, C., & Roodt, G. (Eds). (2001). *An Introduction to Psychological Assessment in the South African Context*. Oxford: Oxford University Press

Fullan, M. G. (1993). *Change Forces: Probing the depths of educational reform*. Bristol, PA: Falm Press.

Gatherole, S. E., & Baddeley, A. D. (1990) A study of young children learning new names. *British Journal of Psychology, 81*, 439-454.

Geva, E. & Farnia, F. (2005). Learning to read in a second language: Research implications and recommendations for services. *Learning disabilities research and practice, 20*: 39-49.

Golberg, H., Paradis, J.,Crago, M. (2008). Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics, 29*, 41-65.

Gottardo et al., A., Collins, P., Baciu, l. & Gebotys, R. (2008). Predictors of Grade 2 Word Reading and Vocabulary Learning from Grade 1 Variables in Spanish-Speaking Children: Similarities and Differences. *Learning Disabilities Research & Practise, 23*, 11-24.

Griffin, P. (2005). Developing a measure of wealth for primary student families in a developing country: comparison of two methods of psychometric calibration. *Studies in Educational Evaluation, 31*, 192-206.

Gregory, R. J. (2007). *Psychological Testing: History, Principles & Applications, 5th Ed.* Minnesota: American Guidance Service.

Haladyna, T. M. (2004). Developing and Validating Multiple-Choice Test Items, 3rd Edition. Mahwah, New Jersey: Lawrence Earlbaum Associates.

Haynes, M. & Baker, I. (1993). American and Chinese readers learning from lexical familiarization in English texts. In T. Huckin, M. Haynes & J. Coady (Eds.), *Second language reading and vocabulary acquisition* (130-152). Norwood, NJ: Ablex.

Henriksen, B. (1999). *Three dimensions of vocabulary development*. Studies in Second Language Acquisition 21, 303–317.

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement & Evaluation in Counselling & Development, 34*, 177-189.

Hill, K. T. & Wigfield, A. (1984). Test Anxiety: A major educational educational problem and what to do about it. *Elementary School Journal, 85*, 105-126.

Holland, B., Goodman, S., & Walkei, J. (1994). *Coaching athletes with intellectual disability.* Australia: Australian Sports Commission.

Holloway, J. (1994). *Language and Intelligence*. London: Macmillan.

Hotiu, A. (2006). *The relationship between Item Difficulty and Discrimination indices in multiple-choice tests in a physical science course.* Unpublished Thesis, University of Florida Atlantic University.

HPCSA .(2002). Retrieved from: http://www.docstoc.com/docs/19874400/The-Health-Professions-Council-of-South-Africa on the 16th of May 2008.

Huitema, B. E. (1980). *The Analysis of covariance and alternatives*. New York: John Wiley & Sons.

Iversen, G. R. & Norpoth, H. (1987). *Analysis of Variance, 2nd Edition*. Carlifonia: Sage Publications.

Jacobson, P. (2001). *Evaluation of a radio programme teaching English as a second language to South African Primary School Learners*, Unpublished Masters Thesis, University of the Witwatersrand: Johannesburg.

Kane, M.T. (1992) An argument-based approach to validity. *Psychological Bulletin, 112*, 527-535.

Kellogg, G. & Howe, M. J. (1971). Using words and pictures in foreign language learning. *Alberta Journal of Educational Research, 17*: 256-276.

Knuver, A.W.M. & Brandsma, H. (1993). Cognitive and Effective outcomes in school effectiveness research. *School Effectiveness and School Improvement, 4*: 369-387.

Kopstein, F. & Roshal, S. (1954). Learning foreign vocabulary from pictures versus words. *American Psychologist, 9*, 407-408.

Krashen, S. (1985). *Language Acquisition and Language Education*. New York: Alemany Press

Kubiszyn, T. & Borich, G. (2003). *Educational Testing and measurement: Classroom application and practice, 7th Ed*. New York: John Wiley and Sons, Inc.

Kunman, A. J. (1998). *Validation in language assessment: selected papers from the 17th Language Testing Research Colloquium.* Long Beach: Antony John Kunnan.

Linde, B.M. (2006). *Managing your money: Understanding Math Operations involving Decimals and Intergers*. Rosen Publishing Group: New York.

Leong, C.B., & Pikulski, J.J. (1990). Incidental learning of word meanings. In J. Zutel and X. McCormick (Eds.), *Thirty-ninth Yearbook of the National Reading Conference*, 231–240. Chicago: National Reading Conference.

Madibeng, M. M. (2007). *English as a medium of instruction: the relationship between motivation and English second language proficiency*. Unpublished Masters Thesis, University of South Africa: Pretoria.

Mandler, G. & Sarason, S. B. (1952). A study of anxiety and learning. *Journal of Abnormal and Social Psychology, 47*: 166-173.

McKeown, M.G., & Curtis, M. E. (1987). *The Nature of Vocabulary Acquisition*. New Jersey: Lawrence Erlbaum Associations, Publishers.

McKeown, M.G. & Curtis, M. E. (1987). *The Nature of Vocabulary Acquisition*. New Jersey: Lawrence Erlbaum Associations, Publishers.

McNamara, T. *Assessment of Second Language Proficiency*. Melbourne: Elservier Ltd.

Maera, P.M. (1996) (a). *The classical research in vocabulary acquisition lognostics tools for vocabulary research*.

Meara PM. (1996) (b). *The third dimension of lexical competence*. Paper presented at the AILA Congress. Jyväskylä.

Meara, PM. (1996). The dimensions of lexical competence. In: G Brown, K Malmkjaer and J Williams (Eds.) *Competence and Performance in Language Learning.* Cambridge: Cambridge University Press.

Messick, J. (1984). *The new design for the National Assessment of Educational Progress*. Presented at the annual meeting of the American Statistical Association, Philadelphia, August 13-17, 1984.

Messick, S. (1989). "Validity." In R. Linn. *Educational Measurement*, 3rd Ed,13-103. New York: American Council on Education/Macmillan.

Messick, S. (1995). Validity of psychological assessment: validation of inferences from person's responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741-749.

Mezynski, K. (1983). Issues concerning the acquisition of knowledge: Effects of vocabulary training on reading comprehension. *Review of Educational Research, 53*, 253–279.

Mislevy, R.J., Steinberg, L.S., & Almond, R.G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1,* 3-67.

Myburgh, O., Poggenpoel, M., & Rensburg, W. (2004). Learner's Experiences of Teaching and Learning in a second or third Language. *Education, 124.*

Naidoo, G. & Potter, C. S. (2005). *Using interactive radio to enhance classroom learning: reaching schools, classrooms, teachers and learners*. Johannesburg: OLSET.

Naidoo, G. & Potter, C. S. (2007). Ethical Issues in Using Interactive Radio in South Africa. *Open Learning: The Journal of Open and Distance Learning, 22*: 159-165.

Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines, 5,* 12–25.

Nation, I. S. P. (2001). *Learning vocabulary in another language.* England: Cambridge University Press.

Nation, I. S.P. (1990). *Teaching and learning vocabulary.* New York: Newbury House.

Nation, P. (2006). *Language Education-Vocabulary*. New Zealand: Elsevier Ltd.

Nieto, S. (2000). *Affirming Diversity: The Socio-political context of multicultural education, 3rd Ed*. New York: Longman.

No Author. (2007). Schools look into build vocabulary to raise scores. *Electronic Education Report, 14,* 4-5 retrieved from http://www.simbainformation.com.

No Author. (2008). *Word Web Pro 5 Online Dictionary*.

Obeegadoo, L. S. (2002). Distance Education and Open Learning in Sub-Saharan Africa: A literature Survey on Policy and Practice. *Working Group on Distance Education and Open Learning.*

OLSET. (1994). *Curriculum Document Number 1. Johannesburg: The OLSET Radio Learning Programme for Educational Development*. Johannesburg.

OLSET. (1994). *Radio Learning Programme for Education Development: Curriculum Document Number 1*. Johannesburg: Open Learning Systems Trust.

Pany, D., Jenkins, J. R., Schreck, J. (1982). Vocabulary Instruction: Effects on word knowledge and reading comprehension. *Learning Disability Quarterly, 5,* 202-215.

Pearson, P. D., Hiebert, E. H. & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly, 42*, 282-296.

Perraton, H., Robinson, B. & Creed, C. (2001). *Teacher education through distance learning: Technology, curriculum, evaluation.* Paris: UNESCO.

Pikulski, J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher, 58*, 510-519.

Pikulski, J. J. & Templeton, S. (2008). Teaching ad Developing Vocabulary: Keys to long-term reading success. *Current Research in Reading Language*. Litho: Houghton Mifflin.

Pinker, S. (1984). *Language Learnability and Language Development.* Cambridge: Harvard

Potter, C. S. & Naidoo, G. (2004). *Using Interactive Radio to Enhance Classroom Learning: Reaching Schools, Classrooms, Teachers and Learners.* Johannesburg: NADEOSA conference, July.

Potter, C.S. & Naidoo, G. (2006). Using Interactive Radio to Enhance Classroom Learning: Reaching Schools, Classrooms, Teachers and Learners. *Distance Education, 27*, 1, 63-86.

Potter, C.S and Naidoo, G. (2009). Evaluating Large-scale Interactive Radio Programmes. *Distance Education, 30*. 1, 117-142.

Qian,D.D. (1998). *Depth of vocabulary knowledge:Assessing its role in adults' reading comprehension in English as a second language.* Unpublished doctoral dissertation, University of Toronto, Ontario, Canada.

Qian, D. D. (1999). Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review, 56,* 282–308.

Qian, D. D. (2002). Investigating the Relationship Between Vocabulary Knowledge and Academic Reading Performance: An Assessment Perspective. *Language Learning* 52, 513–536.

Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal 19*(2), 12–25.

Read, J. (1989). *Towards a deeper assessment of vocabulary knowledge.* Paper presented at the 8th Congress of the International Association of Applied Linguistics (Sydney, New South Wales, Australia, August 16–21, 1987). Washington, DC: ERIC Clearing House on Languages and Linguistics. (ERIC Document Reproduction Service No. ED301048)

Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing, 10,* 355–371.

Read, J. (1995). *Validating the Word Associates Format as a measure of depth of vocabulary knowledge.* Paper presented at the 17th Language Testing Research Colloquium, Long Beach, CA.

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

Read, J. (2007). Second Language vocabulary assessment: current practices and new directions. *International Journal of English Studies, 7*, 105-126.

Reber, A.S. (1985). *Dictionary of Psychology*. New York: Penguin Books.

Reisberg, D. (2001). *Cognition: Exploring the Science of the Mind. Second Edition*. New York: W. W. Norton & Company.

Resnick, D. P., & Resnick, L. (1977). The Nature of Literacy: An Historical Exploration. *Harvard Educational Review, 47*, 370-385.

Richards, J. C. (1976). The role of vocabulary teaching. *TESOL Quarterly, 10,* 77–89.

Robbins, C. & Ehri, L. (1994). Vocabulary intervention for kindergarten students: comparing extended instruction to embedded instruction and incidental exposure. *Learning Disability Quarterly*. Online version 22 March 2007.

Rose, B. & Tunmer, R. (Eds) (1975). *Documents in South African Education*. Johannesburg: Collier-MacMillan.

Rosenthal, R. & Rosnow, R. L. (1991). *Essentials of Behavioural Research: Methods and Data Analysis.* New York: McGraw-Hill.

Rudner, L. M. & Schafer, W. D. (2001). *Reliability*. ERIC Clearinghouse on Assessment and Evaluation. Retrieved from: http://www.ericdigests.org/2002-2/reliability.htm. (ERIC Document Reproduction Service No. ED458213).

South African Netherlands Research Programme in Alternatives in Development. (2008). Retrieved from: www.**sanpad**.org.za on the 8th of August 2008.

South African Institute of Race Relations. (2008). Retrieved from: http://www.sairr.org.za/research-and-publications/south-africa-survey-2008-2009 on 12 August 2009.

Savignon, S. J. (2006). *Communicative Language Teaching*. USA. Elsevier Ltd.

Scheppers, R. (2006). The effects of immersion on Grade 7 learner's vocabulary size: Is incidental learning of vocabulary enough. *Journal for Language Teaching, 40, 1-20*.

Schiefelbusch, R. L. & Pickar, J. (1984). *The Acquisition of Communicative Competence.* Baltimore: University Park Press.

Sénéchal, M., & Cornell, E.H. (1993). Vocabulary acquisition through shared reading experiences. *Reading Research Quarterly*, *28*, 360-374.

Shen, M. (2008). EFL learner's response to extensive reading: Survey and pedagogical applications. *The Reading matrix, 8:* 111-123.

Silva, A.S.F. (2008). *Open Learning, Educational Development and Empowerment: The Case of the Open Learning Systems Education Trust (OLSET).* Unpublished doctoral dissertation. Johannesburg: University of the Witwatersrand,

Spinelli, E. & Siskin, H.J. (1992). Selecting, Presenting and practicing vocabulary in a culturally-authentic context. *Foreign Language Annals*, *25*: 305-315.

Stavonich, K. E., Nathan, R. G., Vala-Rossi, M. (1986). Developmental changes in the cognitive correlates of reading ability and the developmental lag hypothesis. *Reading Research Quarterly, 21*, 267-283.

Surty, E. (2006). *National Consultative Workshop on the Health and Wellness Programme in Education, 24 July, 2006*. Retrieved from: http://www.info.gov.za/speeches/2006/06072410451002.htm.

Swanson, B. B., Payne, D. A. & Jackson, B. (1981). Predictive validity for the Metropolitan. *Educational & Psychological Measurement, 41*, 575-578.

Swanson, E. A. & Howerton, D. (2007). Influence Vocabulary Acquisition for English Language Learners. *Intervention in school and clinic, 42*, p.290-295.

Taylor, A.; Kroch, A.; & Ringe, D. (2000). *The middle English verb-second constraint: A case study in language contact and language change.* Philadelphia: John Benjamins.

Terman, L.M. (1918). *The intelligence of School Children*. New York: Houghton Mifflin Co.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1997). Prevention and remediation of severe reading disabilities: Keeping the end in mind. *Scientific Studies of Reading,* 217-234.

Tran, A. (2007). *A Learning-center Vocabulary-Reading Activity for English-Language.* Oxford: University Press.

Tredoux, C.G. & Durrheim, K.L. (2002). *Numbers, Hypotheses and Conclusions*. Juta: Cape Town.

Ukrainetz, T. A. (2002). The criterion validity of four vocabulary tests compared with a language sample. *Child, language & Therapy, 18*, 59-78.

Van Der Linde, C.H. (2006). The Role of Good Education Management in a changing South Africa. *Education, 122.* 35-47.

Volante, L. (2005). Accountability, student assessment, and the need for a comprehensive approach. *International Electronic Journal for Leardership in Learning, 6,* 34-56.

Wainer, H., Thissen, D., (1989) & Sireci, S.G. (1991). On the reliability of the testlet-based tests. *Journal of Educational Measurement, 28*, 237-247.

Wasik, B. A. (2006). Building vocabulary one word at a time. *Young Children, 61*, 70-77.

Wesche, M. & Paribakht, T.S. (1996). Assessing second language vocabulary knowledge: Depth versus breadth. *The Canadian Modern Language Review, 53* (1), 13-40.

Westerlund, M. & Lagerberg, D. (2008). Expressive vocabulary in 18 month-old children in relation to demographic factors, mother and child characteristics, communication style and shared reading. *Child: Care, health & Development, 34*, 257-266.

Wikipedia Contributors. Retrieved 7 March 2009 from: http://wikipedia.org. Citation. Wikipedia, The free encyclopedia.

Wordweb Online. (2007). *Word Web Dictionary and Thesaurus, 2005-2009*. Wordweb Software.

Wright, N. (1977). Progress in education: A review of schooling in England and Wales. London: Croom Helm Ltd.

APPENDICES

Appendix A

Pilot Standardisation Sample

Name of the school	Number of Grade 1 Learners
Bree Primary School	43
Welizibuko Primary School	34
Paul Mosaka Primary School	43

Appendix B

Background on the Organisation

The OLSET was established in 1992, funded by the USAID. The vision of the organisation is to improve the effectiveness of teaching and learning in impoverished, urban and rural primary school classrooms alike, using innovative Open and Distance Learning Interactive Radio programmes in South Africa and beyond through effective in-service support for Teachers and the provision of appropriate Learner Support Materials. OLSET uses their radio programme called 'English in Action' (EIA) which offers radio lessons to schools, accompanied by print and display materials for teachers and learners in the classroom, to perform the core function of the organisation. The organisation also offers non-prescriptive in-service training and support to teachers using the programme through workshops and class visits. The EIA is used in seven provinces in the country namely; Limpopo, North West, Eastern Cape, KwaZulu Natal, Mpumalanga and the Free State. The OLSET 'English in Action' radio learning programme, serves to provide educational support to marginalized teachers and learners in the country. OLSET has trained 52 000 and reaches 1.8 million learners. It is being used in 11% of all schools in South Africa, and being used by 8% of teachers.

Procedures used by OLSET to get subjects and administer tests

Upon, realising the need to collect this data, OLSET approached 3 schools which have been consistently implementing the EIA programme. These schools we identified by the Programme Regional Co-ordinators and from a wide pool, 3 were randomly selected. OLSET then employed and trained 5 third-year Psychology students from the University of the Witwatersrand in test administration. The students were trained by one Professor from the University, a PhD student who also works full time in the University and a Masters student all from the Psychology Department. An application was sent out to the Department of Education to use these schools for this research and when it was granted, arrangements were then made with the schools to schedule dates and times of test administration. 45 minutes were requested from a school day and the class teachers were involved and assisted in distributing the workbooks and debriefing the learners. Smaller groups were selected from each school the Peabody which is an individual test was administered separately. The learners and schools who participated were thanked.

OLSET has kept all the workbooks with the raw scores.

Appendix C

Metropolitan Readiness Level 1 Test: School Language and Listening Subtest

Concurrent Validity coefficients

Name of test	**r**
Stanford-Binet Vocabulary subtest	0.72
WISC Vocabulary Subtest	0.69
WAIS Vocabulary Subtest	0.64
WPPSI Vocabulary Subtest	0.30
Full-Range Picture Vocabulary Test	0.86
Quick Test	0.72
Van Alstyne Picture Vocabulary Test	0.70
Expressive One-word Picture Vocabulary test	0.70
Pacific Picture Vocabulary	0.50
ITPA Verbal Expression Subtests	0.36

Appendix D

Peabody Picture Vocabulary Test

Split-half Reliability coeffiecient for Form L.

Age group	Number of test takers	r
2years,6months-2years,11months	99	0.67
3 years -3 years,5 months	87	0.78
3 years,6 months -3 years,11 months	105	0.80
4 years -4 years,5 months	116	0.71
4 years,6 months -4 years,11 months	101	0.70
5 years-5 years,5 months	100	0.79
5 years,6 months -5 years,11 months	98	0.73
6 years	94	0.84

Alternate-Forms Reliability Coefficients

Age group	Number of test takers	r
2years,6 months-2years,11months	20	0.83
3 years -3 years,11 months	57	0.82
4 years -4 years, 11 months	63	0.74
5 years.-5 years, 11 months	52	0.80
6 years-6years, 11 months	60	0.78

Appendix E

OLSET Picture-Language Vocabulary Level 1-Test

Appendix F

Item Discrimination Calculations

An example of calculating IDI, with a sample of (N=500), the discrimination index will be calculated as follows; D= U/Nu-L/Ne or simply. This score will An example to illustrate this;

The letters with the asterisk (*) are the correct options in the choice. Higher scorers and Lower scorers refer to the top and low 25% respectively.

Question 1

	A	B	C*	D
Higher Scorers	20	5	80	20
Lower Scorers	25	20	75	5

This is the ideal IDsI where more people in the Upper 25% picked the correct option.

Question 2

	A	B*	C	D
Higher Scorers	15	85	0	25
Lower Scorers	50	25	0	50

This question indicates problems and requires revision as item C is not a good distracter and obviously stands out as the wrong choice for both the High and Low scorers.

Question 3

	A	B	C	D*
Higher Scorers	20	70	15	20
Lower Scorers	5	35	80	5

Item B is problematic and requires revision as more people in the High scoring 25% picked this item more than they did the correct alternative for this which was item D.

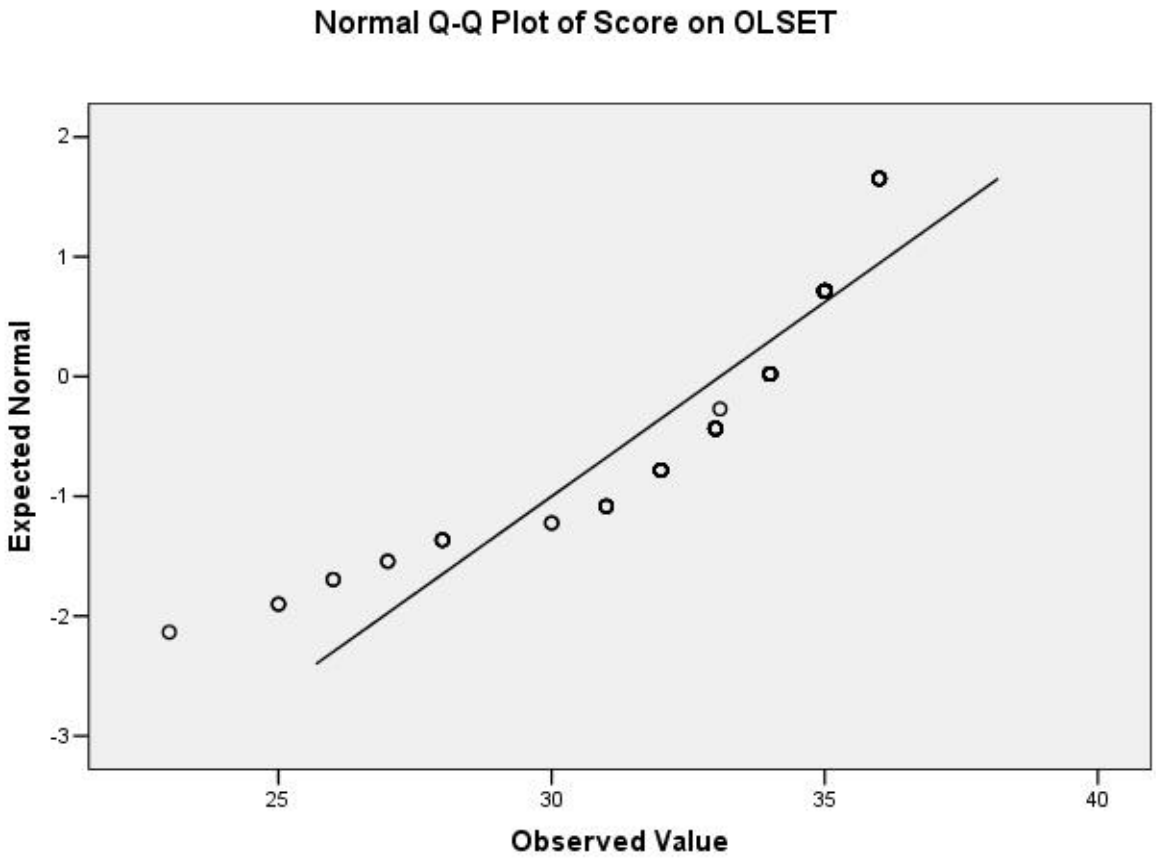
Appendix G

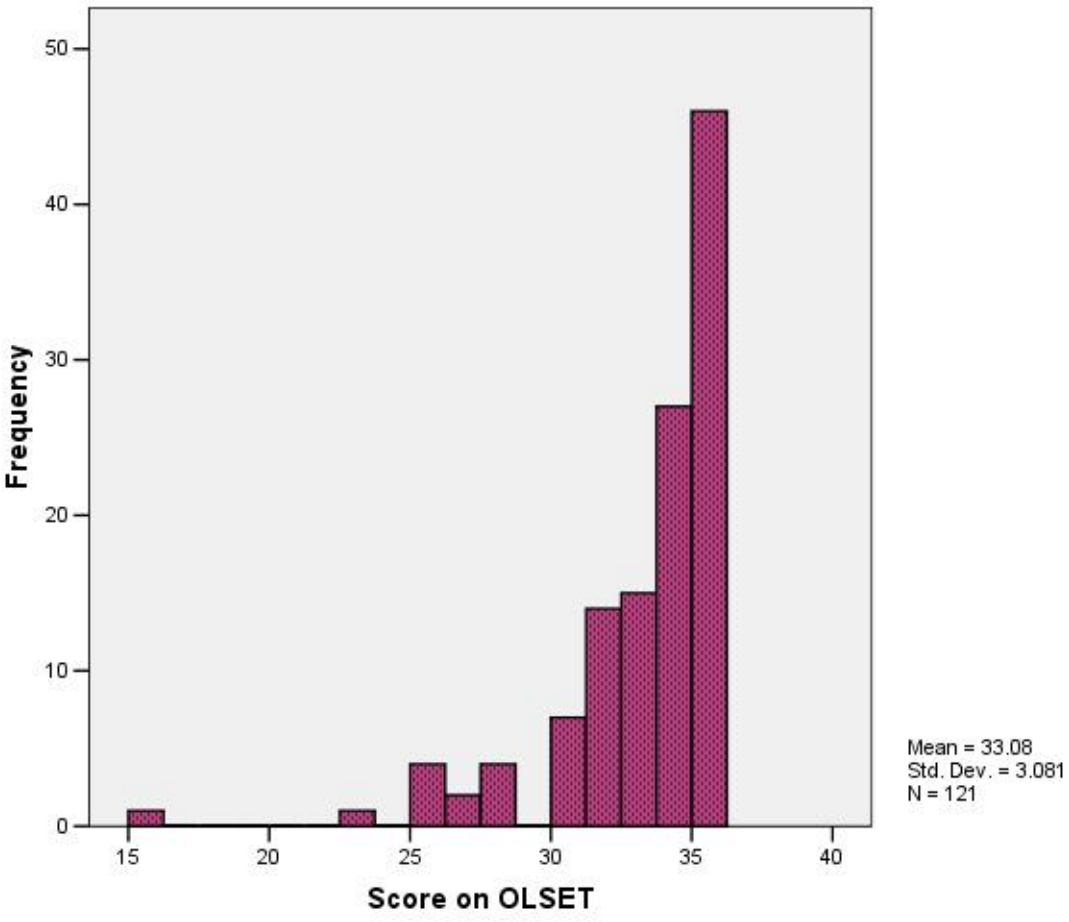
Scores attained by learners speaking different languages

First Language	N	Variance	Mean	Std. dev	Minimum	Maximum
English	5	Metropolitan	14.00	2.00	11.00	16.00
		Peabody	29.60	5.94	20.00	35.00
		Olset	35.20	0.44	35.00	36.00
Malawian	1	Metropolitan	12.00	.	12.00	12.00
		Peabody	23.00	.	23.00	23.00
		Olset	35.00	.	35.00	35.00
Shangaan	4	Metropolitan	7.75	5.79	0	14.00
		Peabody	21.50	5.97	15.00	29.00
		Olset	33.50	2.38	31.00	36.00
Somali	2	Metropolitan	10.50	4.94	7.00	14.00
		Peabody	23.00	12.72	14.00	32.00
		Olset	29.50	6.36	25.00	34.00
Urdu	7	Metropolitan	9.57	2.57	7.00	14.00
		Peabody	22.00	8.96	11.00	34.00
		Olset	34.71	0.48	34.00	35.00
isiXhosa	35	Metropolitan	6.22	3.04	0	13.00
		Peabody	19.17	4.54	9.00	30.00
		Olset	33.40	2.62	25.00	36.00
isiZulu	15	Metropolitan	10.73	3.34	6.00	16.00
		Peabody	24.53	4.15	17.00	31.00
		Olset	32.80	1.32	30.00	34.00
seSotho	36	Metropolitan	8.50	2.97	2.00	16.00
		Peabody	21.22	4.75	12.00	34.00
		Olset	32.66	4.09	15.00	36.00
seTswana	14	Metropolitan	10.07	3.62	5.00	16.00
		Peabody	23.	4.82	15.00	31.00
		Olset	32.42	3.37	23.00	35.00

Appendix H:

Test for Normality of OPLVLT Data: Probability plots




Frequency
50
40
30
20
10
0
15
20
25
30
35
40
Score on OLSET
Mean = 33.08
Std. Dev. = 3.081
N = 121