

President's Education Initiative

DEVELOPMENT OF DIAGNOSTIC TESTS
TO ASSESS GRADE 7 LITERACY AND NUMERACY SKILLS

Submitted to the
Department of Education
and the
Joint Education Trust

9 October 1998

EXECUTIVE SUMMARY

St. Barnabas College set out to develop assessment tools in the key areas of literacy and numeracy that could be used to evaluate learner attainment at a Grade 7 level. Grade 7 was chosen because it constitutes the first year of the Senior phase of education. In the present context, the Grade 7 year is situated in primary schools and achievement in this year of school can be used to direct learners towards appropriate secondary school placements. .

The design of the Maths and English Language assessment tools was informed by Curriculum 2005 and Outcomes Based Education principles - thus criterion-referenced. A primary concern of the research was to determine optimal levels of performance of learners in disadvantaged schools, most of whom are English Second Language speakers.

A first Prototype of the tests was developed and trialed at St. Barnabas College in Gauteng and Bophepha Primary in Wolmaranstad. Based on the learners' performance and the testers' observations, the tests were modified (as Prototype 2) and tested in four "high-performing" primary schools in the Northern Province and Gauteng. Norm-referencing analysis was conducted on Prototype 2 in order to ascertain the relative ease or difficulty of the tests.

Analysis of learner performance on Prototype 2 shows that the tests are fairly good discriminators of Maths skill and performance at the Grade 7 levels. On the other hand, neither the literacy or the numeracy test is challenging enough for higher achieving learners. Although there were significant differences in performance of Grade 6 and Grade 7 learners on the tests, the distribution of Grade 7 scores was generally clustered above the expected mean (the 50th percentile of possible scores) indicating that the tests were not sufficiently challenging or difficult enough for Grade 7 learners. This confirms the opinion of one expert who stated that the tests were not challenging or difficult enough.

The researchers conclude that while the tests require the inclusion of more difficult items, they are valid tools for assessing learner attainment of disadvantaged ESL learners at a Grade 7 level.

TABLE OF CONTENTS

1.	INTRODUCTION	1
2.	BACKGROUND AND RATIONALE.....	1
3.	METHODOLOGY.....	2
3.1	CONSTRUCTION OF TESTS: ALIGNMENT OF TESTS WITH PRINCIPLES OF OBE	3
3.1.1	<i>English Language Test.....</i>	4
3.1.2	<i>Test of Mathematical Ability.....</i>	5
3.2	DEVELOPMENT AND ADMINISTRATION OF PROTOTYPE TEST 1.....	6
3.3	DEVELOPMENT AND ADMINISTRATION OF PROTOTYPE TEST 2.....	7
3.4	DATA ANALYSIS.....	8
4.	ASSUMPTIONS AND DEFINITIONS.....	8
5.	FINDINGS.....	9
5.1	PROTOTYPE TEST 1	9
5.2	PROTOTYPE TEST 2.....	10
5.2.1	<i>General Characteristics of the Sample.....</i>	10
5.2.2	<i>Measures of Test Validity.....</i>	11
5.2.3	<i>Effects of Home Language on English Language Test Performance</i>	19
5.2.4	<i>Performance of the Individual Elements of the Language Test.....</i>	20
5.2.5	<i>Performance of the Individual Elements of the Maths Test.....</i>	25
5.2.6	<i>Equivalency of the Three Test English Language Forms (Tests A, B, and C).....</i>	33
5.2.7	<i>Gender Differences in Performance on Language and Maths Test.....</i>	34
5.3	EXAMINATION OF TESTS IN CURRENT USE AT SAMPLE SCHOOLS.....	36
5.3.1	<i>English Tests in Current Use.....</i>	36
5.3.2	<i>Maths Tests in Current Use.....</i>	36
6.	DISCUSSION.....	36
7.	FUTURE RESEARCH REQUIRED	38

Appendix A: Prototype Test 1

Appendix B: Prototype Test 2 (with forms A, B, and C of the English Language Test)

Appendix C: Examples of Tests in Current Use at Sampled Schools

Appendix D: Profiles of Schools participating in Prototype 2 Tests

Appendix E: Normal Probability Test Plots

Appendix F: Comments from Nick Taylor (Joint Education Trust)

ACRONYMS

C2005	Curriculum 2005
ESL	English as a Second Language
ESP	Education Support Programme
HSRC	Human Sciences Research Council
NGO	Non-Governmental Organisation
NQF	National Qualifications Framework
OBE	Outcome Based Education
SAQA	South African Qualifications Authority
SHC	Sacred Heart College

TEAM MEMBERS

Moira De Groot - Principal Investigator (St. Barnabas College)

Ms. De Groot is an educational consultant with extensive experience in the field of compensatory education for disadvantaged learners and in the development of innovative assessment measures. She has many years of experience in classroom teaching with disadvantaged ESL learners and has been centrally involved in creating assessment tools for the measurement of learner potential at St. Barnabas College and at a number of other schools in Gauteng. She has a Masters degree in Education from the University of the Witwatersrand, Johannesburg, and holds a registration in Educational Psychometry with the South African Medical and Dental Council.

Mary Pat Selvaggio - Director of Research, Khulisa Management Services

Ms. Selvaggio is a public health specialist with extensive experience in managing research activities at the national and community level. She has designed survey instruments, trained fieldworkers in data collection, undertaken extensive data analysis, and prepared numerous research reports for NGOs universities, and government departments in southern Africa. Prior to joining Khulisa, Ms. Selvaggio worked for two years as an independent consultant in southern Africa, and for eleven years as a U.S. Foreign Service Officer responsible for designing and managing health/population assistance programs in Zimbabwe, Mozambique, Swaziland, and other countries in southern Africa. Ms. Selvaggio has a Masters in Public Health from the University of Minnesota, Minneapolis, USA.

Ana Ferreira - Development of English Language Test (St. Barnabas College)

Ms. Ferreira is Head of the English Department at St. Bamabas College. She has extensive experience in the teaching of disadvantaged ESL learners and in assessment of learner potential.

Gary Kelly - Development of Maths Test (St. Barnabas College)

Mr. Kelly is Head of the Maths Department at St. Barnabas College. He has experience in the teaching of disadvantaged ESL learners. He holds a Masters degree in Mathematics from the University of the Witwatersrand, Johannesburg.

June Virtue - Co-ordinator of Logistics (St. Barnabas College)

Ms. Virtue was responsible for the management and administration of the testing conducted at six schools. She is the Director of the St. Barnabas Education Support Programme.

ACKNOWLEDGEMENTS

The team wishes to thank Dr. Peter Fridjhon at the University of Witwatersrand for his assistance in clarifying the statistical tests to be performed on the data. The team also wishes to thank Ms. Celeste de Gouveia at Sacred Heart College (SHC) for her assistance in providing expert opinion on the tests and having the tests administered to SHC learners. Lastly, the team thanks Mr. Nick Taylor of Joint Education Trust for his critical and insightful comments on the draft report (see Appendix F).

1. INTRODUCTION

Under the President's Education Initiative, St. Barnabas College and K hulisa Management Services received funding to develop an assessment tool for evaluating attainment levels of Grade 7 learners. The need for an objective, standardised assessment tool to measure learner attainment is highlighted by the fact that Grade 7 constitutes the first year of the Senior phase of education.

As stated in the proposal, the purpose of this activity was "to develop a diagnostic test that could be used in a systemic evaluation of Grade 7 and assist the State and other training programmes to target their interventions more accurately. Should the pilot prove successful, this assessment tool could contribute to systemic and other evaluations of NGO projects, Curriculum 2005, and other teacher training programmes."

The central role of systematic evaluation of progress in implementing Curriculum 2005 is highlighted in Section 3 of the SAQA document (Technical Committee of SAQA: Towards a Model for Assessment in Terms of an Outcome-based Approach):

"To ensure that the degree and quality of education that is offered by learning sites throughout South Africa maintains standards in accordance with national expectations, a systematic effort must be made to measure the educational process, assess the achieved outcomes and evaluate the findings, so that success can be consolidated and failure remedied." (p.43)

2. BACKGROUND AND RATIONALE

In South Africa there are very few indicators of learner attainment available. Class grades tend to be unreliable and difficult to generalise in an environment where teachers are not always accountable, often do not complete the learning programme and lack content knowledge. Thus, it is difficult to identify the success of new educational programmes or NGO or other teacher training interventions.

The need for appropriate assessment tools for monitoring progress and attainment levels of learners is a central concern as regards the implementation of the new educational curriculum (Curriculum 2005). Concerns regarding subjectivity and lack of consistency in assessment practices are noted in SAQA documents:

"Routine procedures for periodically examining the consistency of assessment judgements need to be established in all learning sites in order to enhance the reliability of assessment procedures. The subjective nature of outcomes-based assessment will require that a moderation model be implemented in order to ensure the maintenance of standards. Education and training practitioners may find it valuable to have access to standardised instruments that they can use periodically to check on their own assessments of learner progress (Technical Committee of SAQA: Towards a Model for Assessment in Terms of an Outcome-based Approach. p. 38).

Given that the Department of Education plans to pilot Curriculum 2005 in Grade 7 in the near future, there is need for a system which will allow quick and easy diagnostic assessment of the success of the programme (at learner level). Such an assessment system will also ensure that implementation of Curriculum 2005 focuses on the most important issues.

St. Barnabas College has considerable experience in assessing educationally disadvantaged English Second Language (ESL) pupils. St. Barnabas has found it necessary to construct its own criterion-referenced tools, because of its experience that local norm-referenced tests (such as those developed by the HSRC) have limited value largely because they have been standardised on non-disadvantaged learner populations. In contrast, the tools developed by St. Barnabas focus on identifying potential among ESL learners from disadvantaged backgrounds. This experience and expertise (in assessing pupils with above average academic potential) was drawn on in this study to develop two new assessment tools in the key areas of English Language and Mathematics.

The need for a system of criterion-referenced assessment has been identified as a key concern of the practice of Continuous Assessment in OBE:

"In a Continuous Assessment-based model, the principle of criterion-referenced assessment underpins all assessment done in the learning site, i.e. measuring individual performance against clearly defined standards"(Technical Committee of SAQA: Towards a Model for Assessment in Terms of an Outcome-based Approach) p. 27).

The Maths and English Language assessment tools developed in this study can serve multiple purposes:

- To describe current attainment levels of Grade 7 learners at a particular school for the purpose of comparing achievement levels at this school with other schools in the area or against a national profile.
- To provide pre-and post-test measures for evaluating the success of intervention/input at a school. Here the parallel or equivalent forms of the tests would be needed, as learners could be assessed on the same areas of behaviour without interference of a test-wise factor.
- To provide pre- and post-test measures over a period of time, for the purpose of assessing the extent to which learner attainment reflects the implementation of OBE principles.
- To diagnose areas of strength and weakness in a particular school. This information could be used to focus intervention programmes at the school.

To diagnose strengths and weaknesses of individual learners in Maths and literacy skills. This could also serve to identify learners with special needs, for whom appropriate recommendations could then be made.

In this project, St. Barnabas College was responsible for developing the instruments, selecting sample schools, and administering the tests in the sample school. Khulisa Management Services was responsible for entering data into the database and statistic analysis. Both parties jointly wrote the draft and final report.

3. METHODOLOGY

The Mathematics and English Language tests created under this study were developed jointly by Ms. De Groot and the Heads of Department (English and Mathematics) of St. Barnabas College.

Development of the two tests occurred in two stages. The first draft (Prototype 1) was developed and tested in Grade 7 classes at St. Barnabas College and at Bophepa Primary in Walmaranstad. Based on the results of this draft, the tests were then revised (thus creating

Prototype 2) and administered in Grade 7 classes in four "high performing" schools in the Northern Province and Gautengl.

High performing, yet still disadvantaged, schools were sought out for the purpose of formulating criteria that would identify optimal levels of performance for disadvantaged ESL learners (rather than simply reflecting low levels of performance of ESL disadvantaged learners). See Appendix D for profiles of each high-performing school, and an explanation of how each school was identified. In each school, a small group of Grade 6 learners were also recruited to take the test for the purpose of serving as a "control" in measuring the tests' validity.

Where Prototype 2 was administered, the schools were requested to provide copies of English and Maths tests in current use at the Grade 7 level (see Appendix C). As will be discussed later in this document, these existing tests fall short in measuring and reflecting the higher cognitive levels consistent with Curriculum 2005.

Finally, expert opinion on the validity of the tests was sought from the Director of Curriculum in a high-performing school in Gauteng. This is discussed in Section 5 below "Findings".

3.1 Construction of Tests: Alignment of Tests with Principles of OBE

The two assessment tools (English Language Test and Mathematics test) were designed to operationalise the relevant SAQA Critical Outcomes and principles of Outcomes Based Education (OBE) at a level appropriate to Grade 7 ESL learners. The OBE principles mentioned below were drawn on to inform the construction of both the English language and Mathematics tests.

As defined in the SAQA Critical Outcomes document, all learners need to demonstrate the ability to:

1. Identify and solve problems and make decisions using critical and creative thinking.
2. Collect, analyse, organise and critically evaluate information.
3. Communicate effectively using visual, mathematical, symbolic and/or language skills in various modes.

Further, OBE directives stipulate the following criteria for assessment:

4. Application and use of knowledge
5. Use of knowledge in real life contexts
6. Meaningful tasks
7. Knowledge, abilities, thinking processes, metacognition and affect.

The tests developed in this study aim to reflect the skills or competencies described above in the areas of literacy and numeracy at a level appropriate to Grade 7 learners. The primary focus in developing the tests was whether or not a learner was competent in "communicating effectively using visual, mathematical and/or language skills" rather than mastery of isolated

¹ Please see Section 4 "Assumptions and Definitions" for further explanation of high performing schools.

splinter skills². Attention was also given to the creation of items and tasks that would reflect

"real-life" or "meaningful" contexts for the learners. Further, attention was given to the inclusion of items requiring "higher-level" and "metacognitive" responses. This report does not provide an "item map" which links each of the test items to the Specific Outcomes for Maths and Literacy under C2005, because the Specific Outcomes for the Intermediate Phase (which includes Grade 7) were not prepared at the time of constructing the test. However, as is recommended in Section 7 below, this "item map" needs to be developed in order to assure that the tests address each of the specific outcomes.

Norm-referencing of the data set was undertaken in the analysis to determine the assessments' relative ease or difficulty for disadvantaged ESL learners. This analysis is further explored in Section 5 below.

3.1.1 English Language Test

The English language test developed in this study comprises three sections: Reading comprehension (Lang1), Visual Literacy (Lang2), and Writing skills (Lang3). A total score for the English Test was computed as the sum of the three sections. Three equivalent forms

of the English Test (Test A, Test B, and Test C) were constructed to allow for testing and re-testing without the interference of a 'test-wise' factor (for purposes of establishing reliability).

Reading comprehension skills are assessed in the Lang1 component of the English Language test. In accordance with OBE principles, the following key concerns were identified:

- Identifying messages in texts.
- Making inferences from texts.
- Reflecting critically on the writer's point of view.

In accordance with these principles, questions reflecting three different levels of cognitive skill are used in the Lang 1 section. These levels are:

- lower order questions requiring a literal level of understanding, answers to be found "in the lines"
- middle order questions requiring inferential thinking, answers to be found "between the lines" (i.e. hidden meanings)
- higher order questions (answers are "beyond the lines") requiring the reader to respond critically to the text.

An analysis of the tests in current use show that middle and higher order questions are largely neglected in schools at the present time (see Appendix C).

A multiple-choice answer format for Lang1 was chosen in order to avoid the conflating of receptive and expressive language skills. A cloze item based in the context of the reading comprehension exercise was included to allow for the assessment of reading comprehension skills drawing on syntactic and pragmatic cues.

As can be seen from Appendix C, most of the schools' current tests focus on assessing isolated splinter skills lacking contextual focus.

The design of the Visual Literacy component of the Language Test (Lang2) was informed by the following extract from Curriculum 2005 documentation:

"Initially literacy is seen as a cognitive process that enables reading, writing and numeracy. Currently the use of the term literacy has expanded to include several kinds of literacies. The concept of `literacies' stresses the issue of access to the world and to knowledge through development of multiple capacities and the need to make sense of our worlds through whatever means we have, not only texts and books." (Curriculum 2005 documentation: "Language, Literacy and Communication" p. 25)

In this context, examples of kinds of literacies mentioned are:

- Critical literacy: the ability to respond critically to the intentions, contents and possible effects of messages and texts on the reader.
- Visual literacy: the interpretation of images, signs, pictures and non-verbal language.

In accordance with these principles, the English language test extends the traditional notion of a reading comprehension exercise to include a visual literacy task. This item further allows for the assessment of a learner's ability to respond effectively to a situation depicting a "real-life" problem.

With regard to Writing Skills, OBE criteria stress competence in making meaning as opposed to a more mechanistic emphasis on grammatical correctness. The Lang3 section provides a stimulus for free expression. A learner's writing is then assessed according to the following OBE criteria:

- Structural organisation
- Clarity of expression
- Originality of ideas
- Appropriate use of language (e.g. vocabulary, grammar, spelling, syntax, punctuation)
- Competence in sustaining discourse

Here again, the emphasis on the holistic concept of making meaning is very different to the stress on splinter skills evidenced in the tests collected from the sample schools (Appendix C).

3.1.2 Test of Mathematical Ability

The Maths test developed in this study was designed in four components:

Maths1 - Calculation, computation: This component assesses a learner's ability to perform the four basic operations in Maths, at a level appropriate to Grade 7. OBE Specific Outcome #1 for the area Mathematical Literacy and Mathematics is operationalised in this component of the test: "Learners will be able to demonstrate understanding about ways of working with numbers." (p.109) It is acknowledged that the "Specific Outcomes" mentioned above do not provide specific guidelines regarding the type of numbers to be used in such computations at a Grade 7 level or the appropriate levels of complexity for each respective type of number at this level. The expertise of St. Bamabas Maths teachers was drawn on to develop test items considered to reflect the appropriate level of mathematical ability at a Grade 7 level. The

systematic and comprehensive learning programmes in use at St. Barnabas College were referred to, as a guideline for this process.

Maths 2 - Concepts underlying the four operations, such as magnitude, greater than or less than, ordinal value: This component assesses a learner's ability to use the language of mathematics - which is a key indicator of a learner's understanding of mathematical concepts. This section reflects OBE Specific Outcome #9: "Learners will be able to use mathematical language to communicate mathematical ideas, concepts and generalisations." (p.110)

Maths3 - Spatial reasoning and understanding: This component reflects OBE Specific Outcome #7 for Mathematical Literacy and Mathematics: "Learners will be able to describe and represent experiences with shape, space, time and motion" (p. 110).

Maths4 - Deductive thinking: This component assesses a learner's ability to use the basic operations of mathematics to solve problems: This section reflects OBE Specific Outcome #2: "Learners will be able to manipulate number patterns in different ways" (p.110) as well as the SAQA Critical Outcome relating to identifying and solving problems.

A total score for the Maths tests was computed by summing the individual scores of each of the sections. In contrast to the Language test, only one form of the Maths test was constructed at this stage as it was assumed that parallel forms of the test could be set up at a later stage by simply varying the numbers in each of the mathematical problems.

An analysis of the Maths tests currently used at the sample schools (see Appendix C) show that Maths 1 and Maths4 are usually covered, but that Maths2 and Maths3 are largely neglected. Moreover, present data suggests that Spatial Reasoning and Understanding (Maths3) are seldom covered at any school in South Africa, not even at the high-performing schools selected in this study.

3.2 Development and Administration of Prototype Test 1

Prototype 1, the first draft of the three language sections and four Maths sections, was constructed in November 1997.

Prototype 1 was first tested on the Grade 7 class at St. Bamabas College, Bosmont,

Johannesburg on 2 February 1998. These learners are regarded as representing a high-performing group as they are drawn from disadvantaged schools and impoverished home backgrounds and are recruited on the basis of superior performance on the St. Barnabas Entrance Assessment Tests (an extensive battery of norm-referenced and criterion-referenced tools which over the years have proven to be valid indicators of academic potential among ESL disadvantaged learners). Thus, given their solid performance on the Entrance Assessment Tests, and subsequent admission into St. Bamabas, it was determined that these learners constitute a "high-performing" sample.

Pilot testing of Prototype 1 was conducted very early in the school year in order to avoid "interference" from the enriched educational programme at St. Barnabas. Pupils' performance during the Prototype 1 testing session was closely observed and pupils were encouraged to ask questions, especially with reference to any questions which they experienced as ambiguous or unclear.

A second round of testing of Prototype 1 was conducted at Bophepha Primary in

Wolmaranstad on 26 March 1998. This school was chosen as a balance to the high-performing sample at St. Bamabas, as it was known to be a sound ex-D.E.T. school with a

more "typical" range of learner ability. Performance during the testing session was closely observed and all questions asked by learners were noted for purposes of revising the test.

As described in *Section 5.1*, the tests were revised in the light of findings from the two Prototype 1 testing sessions and a final version of the test (Prototype 2) was constructed in April 1998.

3.3 Development and Administration of Prototype Test 2

Following the revision of the test, Prototype 2 was then administered to four additional high-performing schools in the Northern and Gauteng Provinces. High-performing schools were

identified by asking high-performing senior schools (as defined by the Sunday Times Top 100 schools) to identify those primary schools which serve as feeders for their learner populations.

The first site for the administration of Prototype 2 was Baropodi Primary School, located in the district of Jane Furse in the Northern Province. Testing was conducted on 23 April 1998. The second site for the administration of Prototype B was Little Bedfordview School, located in the district of Lobowakgomo in the Northern Province. Testing was conducted on 21 May 1998. The third site for the administration of Prototype B was St. Scholastica in Venda. Testing was conducted on 18 June 1998.

For the fourth site, arrangements had been made to conduct testing at Kgwadu primary school, Northern Province on 12 June 1998. Unfortunately, the SADTU labour dispute at this time disrupted schooling, so this testing session had to be cancelled. Alternative arrangements were then made to conduct testing at St. Theresa's in Coronationville, Johannesburg. Testing was conducted on 24 June 1998. Socio-economic profiles of each school included in the Prototype 2 sampling are found in Appendix D.

At each site learners were divided into 3 groups of 40 children each. Teachers at the schools had been requested to arrange for 30 Grade 7 and 10 Grade 6 pupils³ in each of the three classrooms. This request was not always strictly adhered to, thus numbers of pupils in the two groups vary somewhat from school to school. At two of the sites (St. Scholastica and St. Theresa's), the entire Grade 7 cohort was assessed as the number of pupils at this level was below 90. One of the 3 different forms of the English Test (Test A,B, C) was administered in each of the classrooms. The same Maths test was administered in all 3 classrooms. Three St. Bamabas testers, who had been fully briefed, administered the tests. The testers were assisted by Grade 7 class teachers in each of the schools.

Pupils were assisted by the testers and by their own teachers to complete the "Cover Sheet" reflecting biographical information. It was found that many pupils were not sure of their date of birth and thus had difficulty completing this section of the Cover Sheet. With the English Test, testers explained each section and went through the *Practice Items* together with the pupils. Specific attention was paid to ensuring pupils had a clear understanding of items and response-choice formats which were not familiar to them. Although pupils were encouraged to ask questions during the course of the testing, it was found that very few questions were asked. Where questions were asked or in instances where pupils appeared uncertain, teachers from the schools were requested to convey information in the predominant language of the school.

³ 10 Grade 6 pupils at each site participated in the testing to serve as a "control" group.

One of the objectives of this study is to provide feedback to all schools participating in the study. This aspect of the project is still to take place. Feedback will be based on data emerging from statistical analysis. Schools will be provided with information regarding the performance levels of their Grade 7 learners as compared to the other schools in the sample. Diagnostic feedback regarding areas of strength and weakness for each school will also be provided. Further diagnostic information regarding strengths and weaknesses of individual learners will assist in identifying learners with special needs.

Follow-up, in terms of teacher development workshops aimed at specific issues, will then be undertaken by St. Barnabas Education Support Programme (ESP) or by other NGOs working with the sample schools.

3.4 Data Analysis

Each test (and its components) was analysed by looking first at the distribution of scores to determine whether the scores were normally distributed around the mean (see *Assumptions and Definitions* below). This analysis was undertaken by utilising the Anderson-Darling normality plot test (see Appendix F) which graphically depicts whether the distribution is normal by calculating a trend line, and which also provides a p-value to indicate the level of significance of the normal distribution.

Each test and its components were also analysed to determine if the actual mean score (and corresponding distribution) was significantly different than the expected mean (the 50th percentile of the possible range of scores for the test or component). This was done by using a 1-sample t-test to ascertain significant differences between the actual mean and the expected mean. If the actual mean was lower than expected (i.e. it was to the left of the 50th percentile), then the test (or component) was deemed to be too difficult. Likewise, if the actual mean was higher than expected (to the right of the 50th percentile) then the test (or component) was interpreted as being too easy.

4. ASSUMPTIONS AND DEFINITIONS

High Performing Learners/Schools: High performing ESL learners were sought out in this study in order to establish a standard of achievement which would be high enough to discriminate excellent achievement from average or poor achievement, but not too high to be discouraging to average-performing learners. If a "typical" learner population were sought out instead of a "high-performing" population, then the resulting assessment tools would have been too low to discriminate between average and high performers. Selection of sites was limited to disadvantaged ESL school populations in order to provide a valid representation of the majority of primary school pupils in South Africa.

As indicated in Section 3.3 of this report, the "high-performing" Grade 7 classes were obtained by asking recognised high-performing high schools in the Northern Province for

information on their "feeder" schools. For example, the principal of a recognised "high-performing" high school was asked to identify primary school(s) which serve as an important feeder school (by virtue of experience that pupils from those schools consistently demonstrated higher achievement levels). Where a single primary school received several recommendations, then that primary school was considered to be "high-performing" relative to other primary schools in the same area and thus selected to be in the sample. Further details regarding the four schools included in the sample is found in Appendix D.

Grade 6 Control Group: Grade 6 ESL learners were included in the sample in order to confirm whether the assessment tools indeed measure Grade 7 levels of achievement. The rationale for this was to provide a "control" group within the same context for the research. The researchers hypothesised that the Grade 6 learners would underperform the Grade 7 learners.

Normal Distribution of Scores: A basic assumption of the analysis below is that a normal distribution of scores (i.e. a "bell-shaped curve") suggests test validity in discriminating between poor, average, or good performance. If the scores of the test (or parts thereof) do not show a normal distribution, then the test (or part) is not sensitive or specific enough to distinguish true learner achievement at the higher and lower levels.

Clustering of the scores around the expected mean (the 50th percentile score of the possible range of scores) was interpreted to mean that the assessment was just the right difficulty to discriminate between poor, average, and high achieving learners. Any significant difference between the actual mean and the expected mean is interpreted to mean that the test (or component) was too difficult or too easy.

Levels of Significance: For the purposes of this analysis in this report, a 95% confidence level is used ($p < .05$). Statistical significance is represented by "****" in this report.

5. FINDINGS

5.1 Prototype Test I

The St. Bamabas sample consisted of 40 Grade 7 pupils, 18 male and 22 female. Ages of these pupils ranged from 10 years, 8 months to 13 years 3 months. Eighty percent of the sample (32 pupils) were ESL speakers (first Languages were Sesotho, Setswana, Venda, Xhosa, Pedi, Zulu or Afrikaans) and 8 were English First Language speakers. Twelve pupils completed Test A of the English Language Test, 15 pupils completed Test B and 13 pupils completed Test C of the English Language tests. All pupils completed the one Mathematics Test.

Time taken to complete the tests averaged 45 minutes, with all pupils being able to finish within 60 minutes.

Despite the efforts to make the three English Language tests equivalent, analysis of scores on the English language test indicated that scores on Test B were highest while scores on Test C were lowest. Item analysis of individual questions suggested that too many Test B questions were "first order" questions, while very few questions reflected "third order" cognitive levels. Item analysis of questions on Test C indicated that a number of questions were ambiguous.

Analysis of performance on the Maths test indicated that some questions required revision. In particular, the instructions for many items needed re-wording in order to be accessible to Second Language speakers.

The second pilot-testing of Prototype I was conducted at Bophepha Primary in Wolmaranstad on 26 March 1998. The principal of the school requested that all Grade 7 pupils in the school be included in the assessment sample, resulting in a total of 208 pupils, 105 male and 103 female. The youngest pupil in the sample was 11 years old, while a number of pupils in the group were 21 years of age. The average age of the group was 15 years 06 months. All pupils were ESL speakers. The First Language of the majority of the pupils was Tswana.

Results of this assessment are not be presented in this report as the attainment levels of these learners was not as significant to the project as the process observed during the testing session. These observations revealed that the Bophepha pupils experienced difficulty with a number of items which the St. Barnabas pupils had found relatively easy. Also, it was found that the time required by this sample to complete the tests was significantly longer than the time taken by the St. Barnabas sample.

All tests were revised in the light of the findings from the two Prototype 1 testing sessions. On the English Language test, more "higher order" comprehension questions were introduced for Test B. A number of questions on Test C were reworded, while others were replaced with less ambiguous questions. The Visual Literacy stimuli were modified to provide fewer distracting factors. The Maths test was revised considerably to provide a more accessible format and fewer language demands. Many items were re-com contextualised to provide a context that would be more meaningful to the target population.

5.2 Prototype Test 2

The revised test (Prototype 2) was then administered to four "high performing" schools in Gauteng and Northern Province.

The findings presented below (sections 5.2.1 through 5.2.6) are for Grade 7 learners only, except where comparisons are made between the Grade 6 "control" and the Grade 7 "experimental" groups.

5.2.1 General Characteristics of the Sample

As shown in Figure 1 below, there were approximately equal numbers of learners between the four schools selected in the sample. Grade 6 learners (the control group) comprised approximately 35 percent of the total sample (N=152) and Grade 7 learners comprised 65 percent of the total sample (N=278).

Figure 1

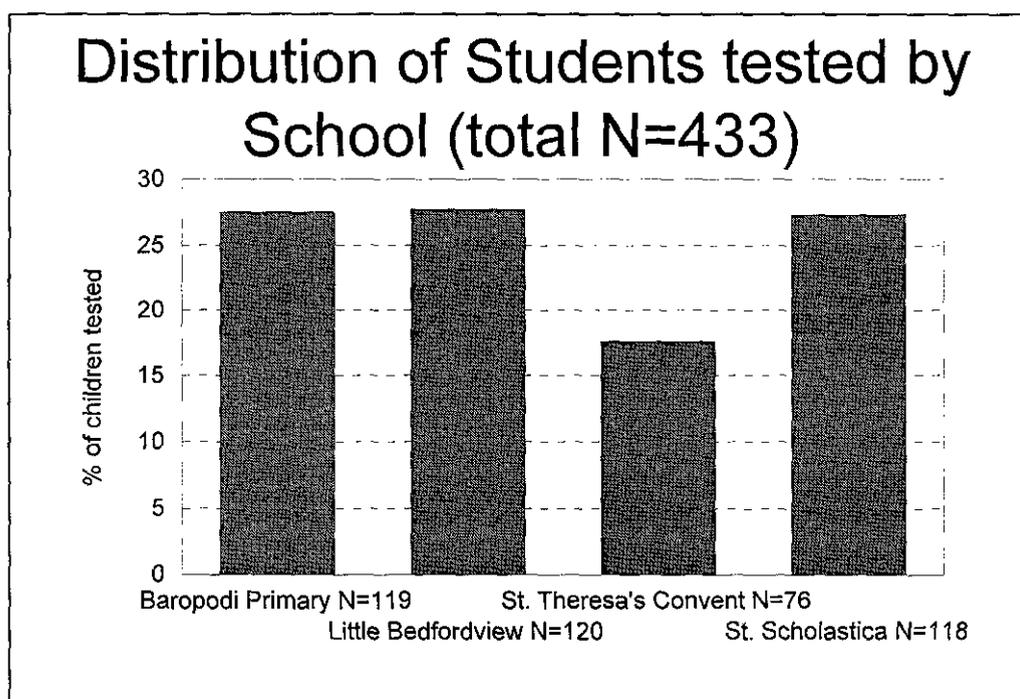


Figure 2

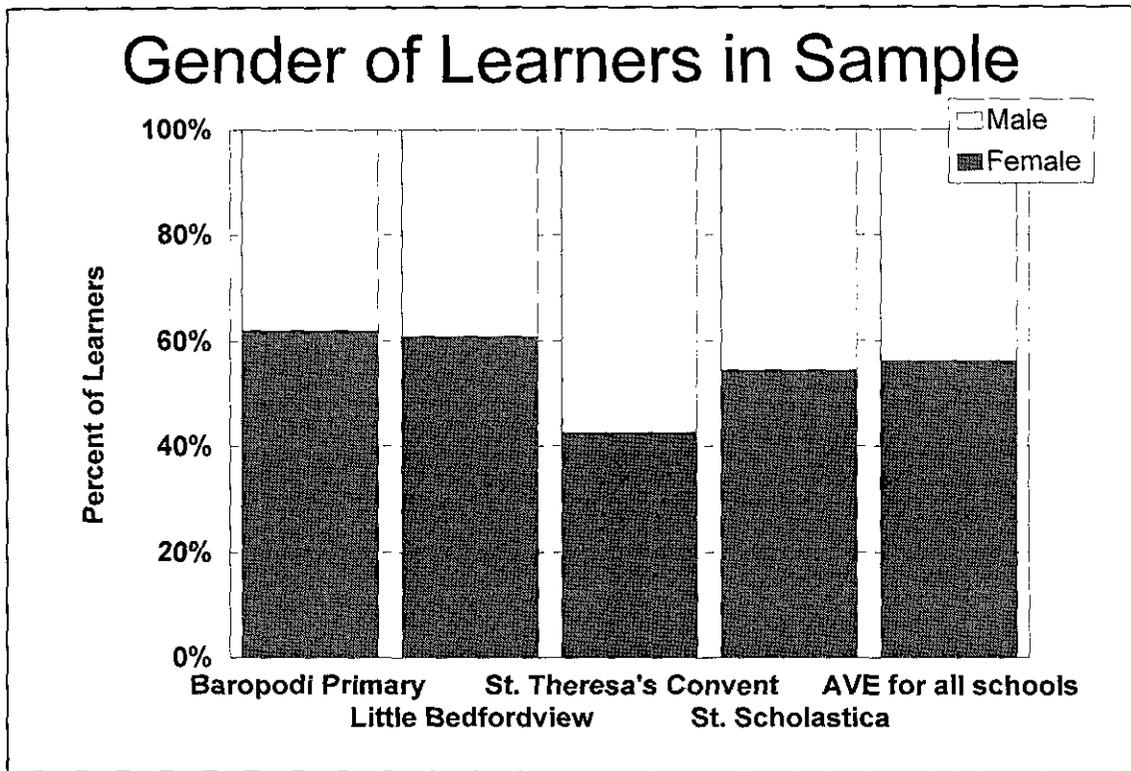
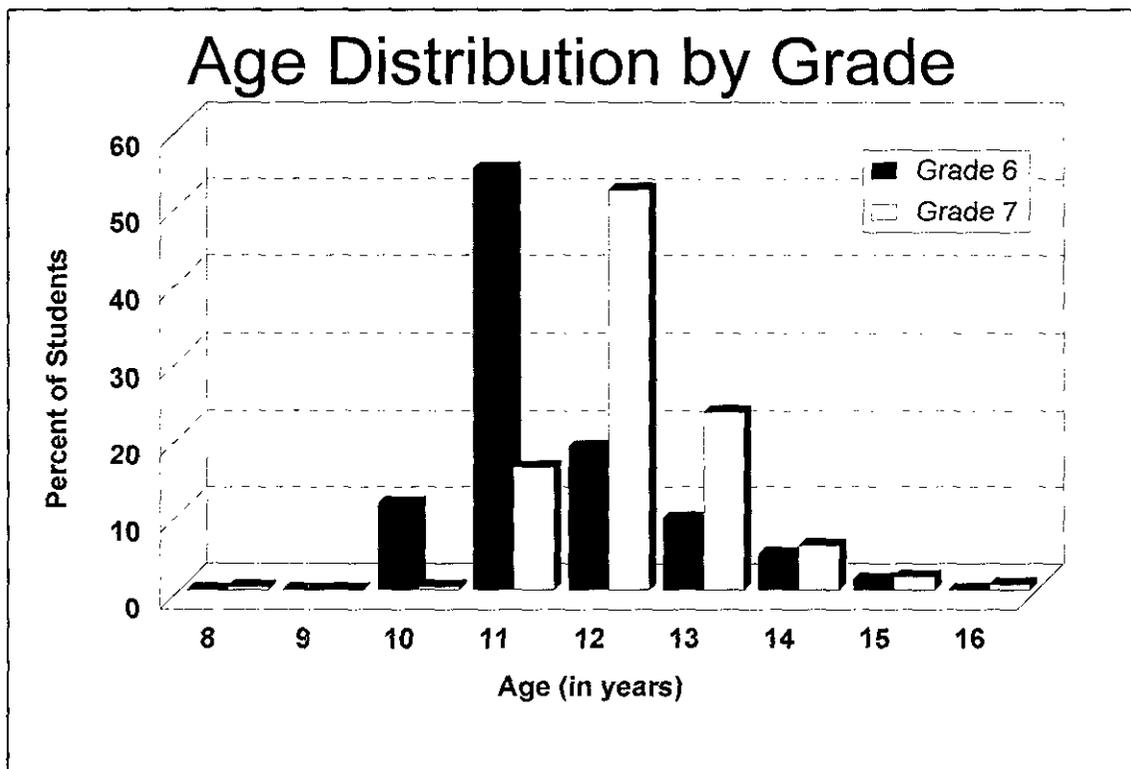


Figure 3



Generally there were more girls than boys in each of the classes selected in the sample (see Figure 2) with an overall average ratio of 6 girls to 4 boys.

As shown in Figure 3 there was a wide range of ages found in the sample, with the youngest age 8 (in Grade 6) and the oldest age 16 (in Grade 7). However, as expected, the average age for the Grade 6 control group was 11 and 12 for the Grade 7 learners. These average ages can be regarded as "appropriate" for Grade 6 and 7 levels. It should be noted that approximately appropriate age levels are important for effective instruction to take place in the classroom, and is likely to be one of the factors contributing to the status of the selected schools as "high-performing". Typically, in schools where there are very wide age ranges in each grade, there is less chance of learners achieving optimal levels of attainment.

5.2.2 Measures of Test Validity

Validity is a measure of the extent to which a test measures what it purports to measure - i.e. how well does the test do the job it is designed to do⁴. It is universally acknowledged that validity is a crucial consideration in evaluating tests and test application.

For the purposes of this study, validity is ascertained through determining (i) the extent to which the tests are consistent with Curriculum 2005 principles (content validity); (ii) the extent to which the distribution of scores discriminates between high, average, and poor performing Grade 7 learners (norm-referencing); (iii) the extent to which the test discriminates between Grade 6 and Grade 7 learners (norm-referencing); and (iv) expert opinion on the content of the tests.

Content validity has been achieved through the construction of the tests. As described in Methodology (Sections 3.1 through 3.1.2), construction of the tests was informed by the principles of OBE and SAQA Critical Outcomes. The content of the tests reflect the end-goal towards which educational practice at a Grade 7 level should be moving, rather than simply capturing a measure of current educational practice at this level. The differences between current practice and 'ideal' practices which accord with the tenets of OBE are revealed in the comparison between the St. Barnabas tests and the assessment tools in current use at the schools used in the sample (see Section 5.3 below).

Normal distribution of scores: The scores for Grade 7 learners for both the language and the Maths tests are normally distributed (see Figure 4 and Figure 5). This is confirmed by normal probability plot tests⁵ that were run on each data set and show significant p-values for the normal distribution of scores for both Language Total and Maths Total.

Further, when the "fit" of the distribution over the expected mean (the 50th percentile of the range of scores) is examined, it can be seen that overall, Language Total scores are distributed as expected (no significant difference from the expected mean $p=.33$), while the distribution of Maths total scores below the 50th percentile (significantly different than expected: $p=.000$) indicates that the test was on the difficult side.

See Gray, B. Thomas, "Controversies Regarding the Nature of Score Validity: Still Crazy after All these Years." Paper presented at the annual meeting of the Southwest Educational Research Association. Austin, Texas. January 1997. (<http://ericae.net/ft/tamu/valid.htm>)

⁵ The normal probability plot test determines if the underlying error distribution is normal or not. When the data set is normal, the plot test resembles a straight line. See Appendix E for the normal probability plots for various data sets presented as graphs throughout this report.

Figure 4

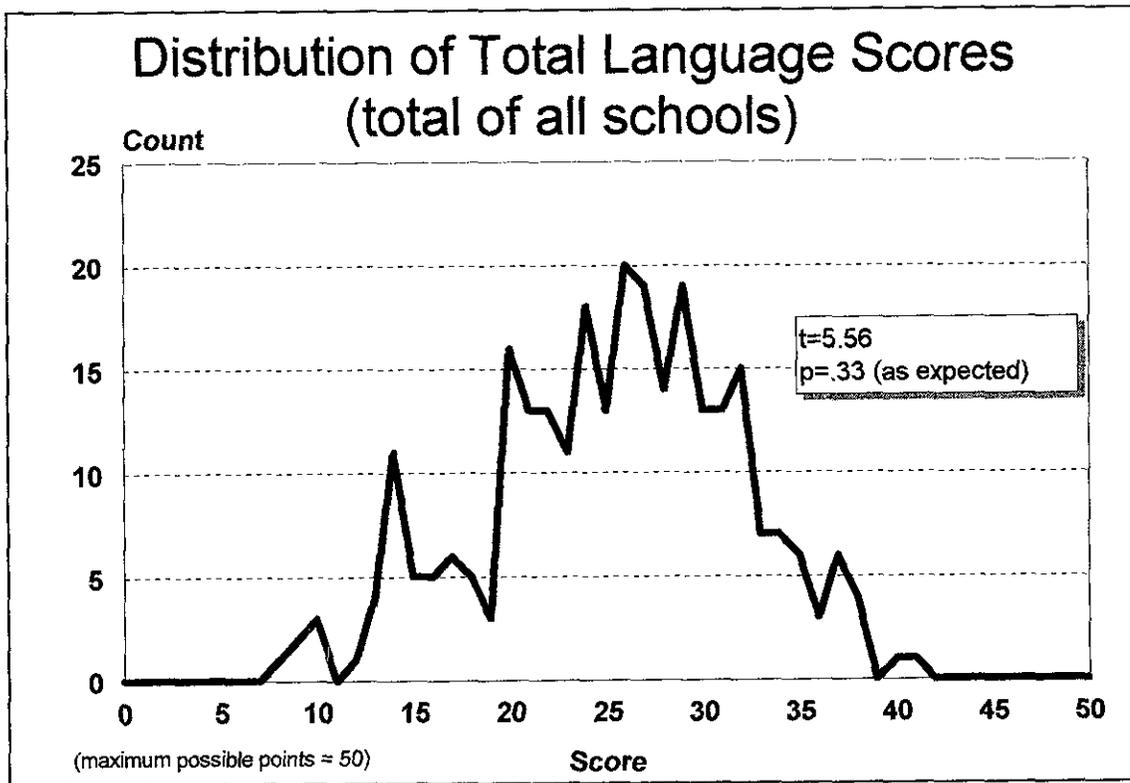


Figure 5

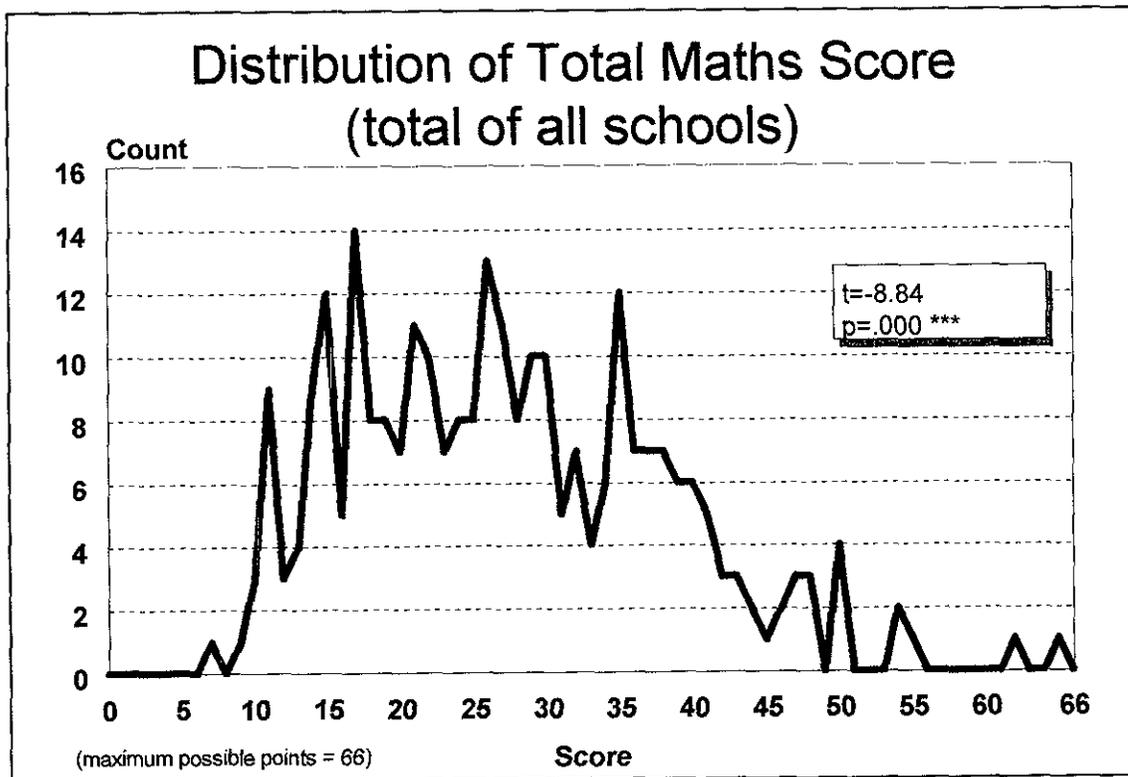


Figure 6

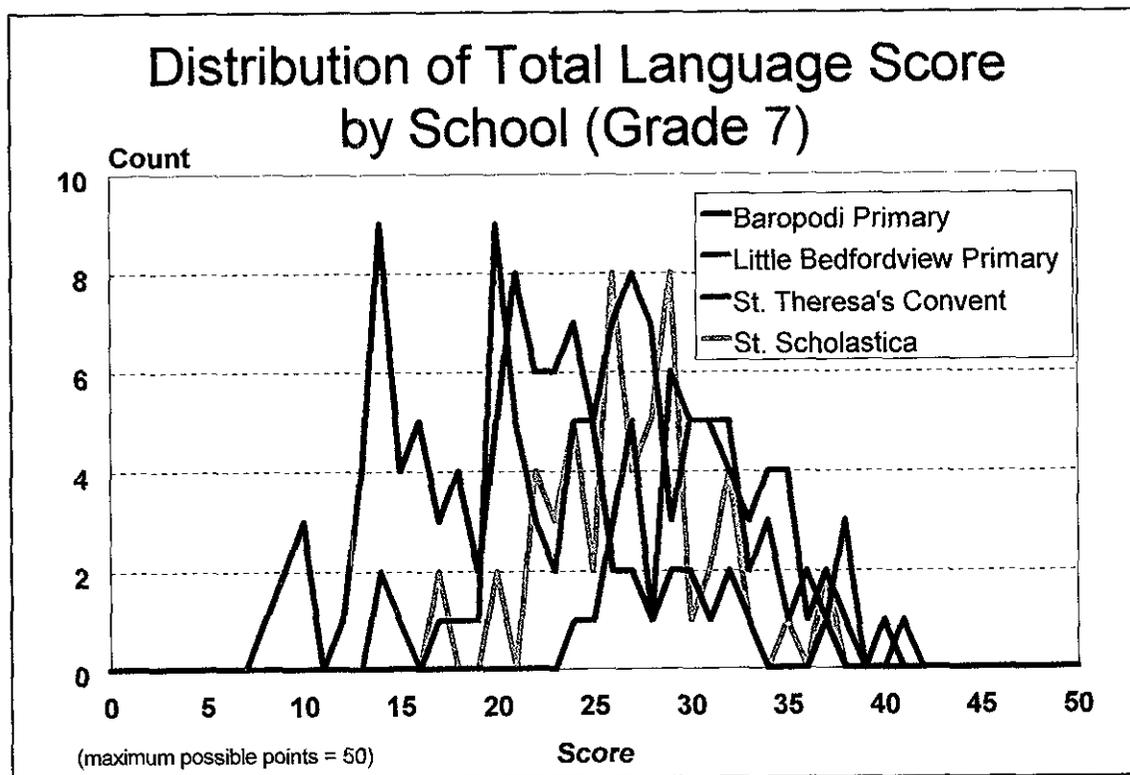


Figure 7

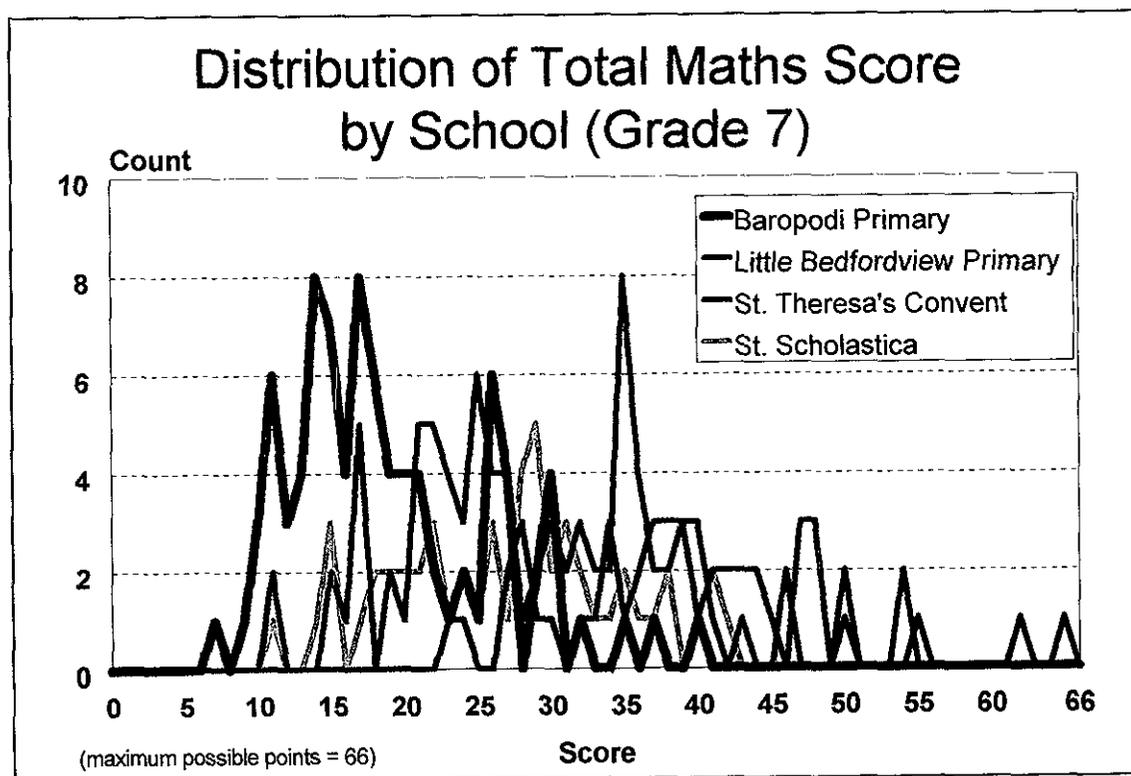


Figure 8

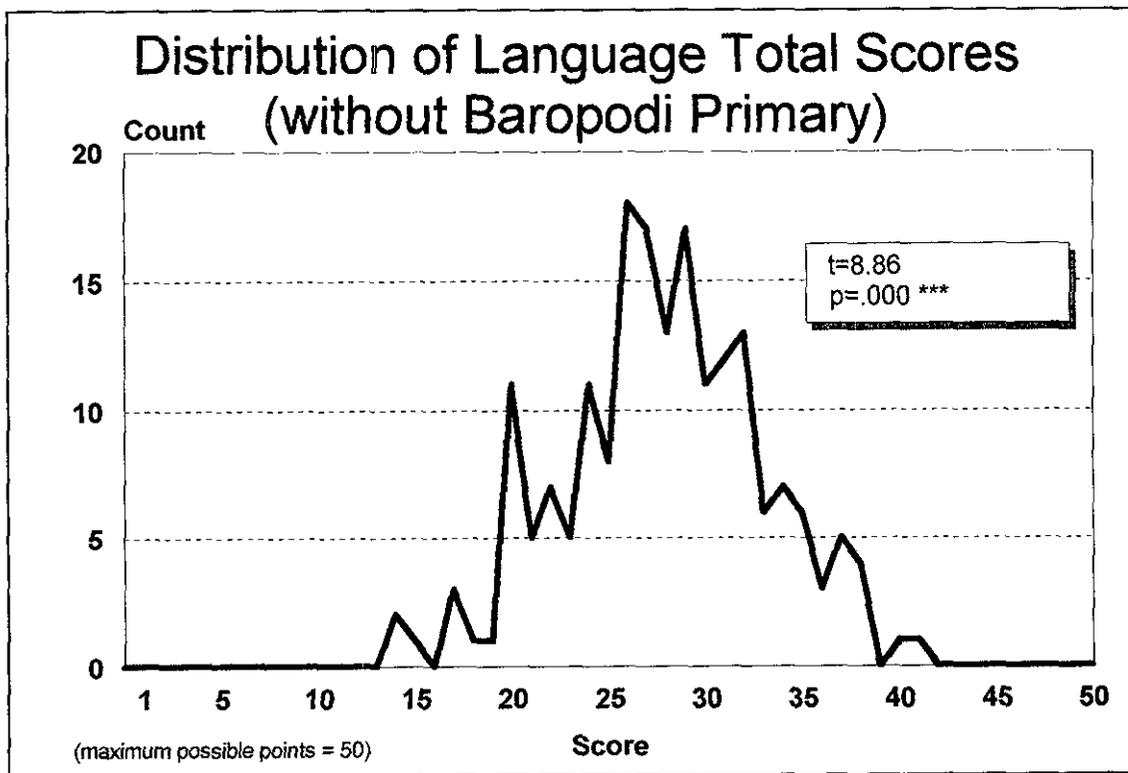
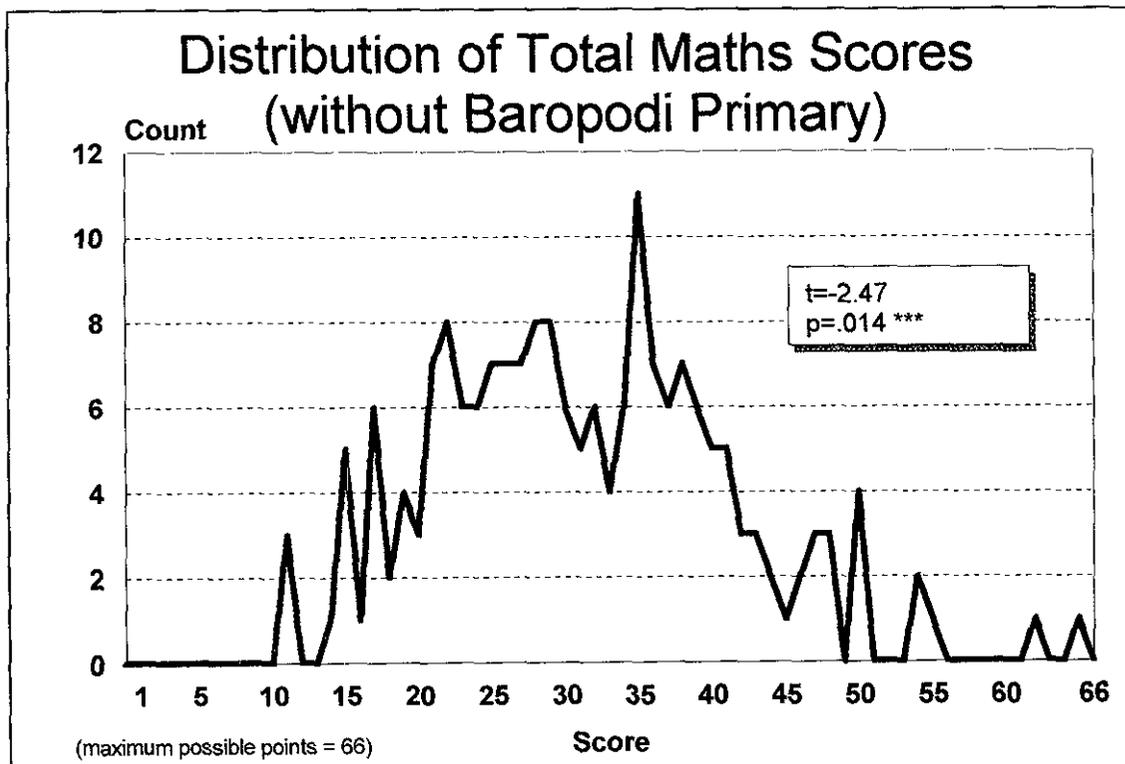


Figure 9



However, when the distribution of scores by school is examined, one school (Baropodi Primary) consistently underperformed (see Figure 6 and Figure 7) and did not exhibit the same "fit" with the expected distribution around the expected mean as compared to the other schools. When Baropodi Primary is removed from the analysis (Figure 8 and Figure 9) it becomes clear that in fact, the Language test is probably too easy (i.e. because the distribution of scores is significantly higher than expected -- around the 60th percentile $p=.000$) and that the Maths test is largely appropriate but perhaps still a bit too difficult (significant difference between the actual and expected mean $p=.014$). These findings are further explored below.

Comparisons in performance between Grade 6 and Grade 7 learners: When the distribution of scores of Grade 6 learners is compared to Grade 7 learners, we see that there is a

significant difference in performance in the English language test between the two grades (2-tailed t-test $p=.017$) (see Figure 10). Likewise, the Maths test reflects even greater significant differences between the levels of performance of Grade 6 compared to Grade 7 learners (2-tailed t-test $p=.0019$) (see Figure 11). This means that the tests are discriminating in measuring performance differences between Grades 6 and Grades 7 learners.

Expert Opinion: Ms. Celeste De Gouveia, Vice-Principal and Director of Curriculum of Sacred Heart Primary (SHC) (Johannesburg), was asked to provide an opinion on the validity of the two tests. SHC Primary is the primary feeder school for SHC High School, recognised as one of the top 100 schools in the Sunday Times survey, and thus considered to be a "high-performing" primary school. Ms. De Gouveia has extensive experience in the teaching of

ESL pupils and holds a Masters degree in Education from the University of the Witwatersrand, Johannesburg.

Ms. De Gouveia scrutinised the English Language test, and it was her opinion that the test was not appropriate to a Grade 7 level with respect to SHC standards. For this reason, the tests were administered to Grade 6 pupils at SHC during July 1998. Discussions with Ms. De Gouveia and three SHC Grade 6 teachers were then held to solicit their opinions on the validity of the tests.

Indeed, on the English Language Test, the Grade 6 learners at Sacred Heart outperformed all the other Grade 7 "high-performing" classes in the sample (see Figure 12). The distribution of scores is focused on the 80th percentile of the test ($p=.000$), indicating that, by Sacred Heart standards, the test is probably not a good discriminator of English Language capacities at the Grade 7 level.

Further, a number of points were raised by the Grade 6 teachers. First, the reading comprehension section (Lang 1) was not perceived to be a good discriminator, particularly in Tests B and C. These tests were found to be too simple and were estimated as reflecting a Grade 5 level by SHC standards. Test A/LangI was found more challenging, but still reflective only of a Grade 6 level. The cloze section of LangI was found to be an accurate discriminator at a Grade 6 level in that this section of the test distinguished pupils of different ability levels. Second, the Visual Literacy section of the test (Lang2) was perceived as appropriate at a Grade 6 level. Third, the free writing section of the test (Lang3) was perceived as the best section of the test, in terms of providing an accurate indicator of a

⁶ Given these results, Khulisa and St. Barnabas conclude that Baropodi Primary was not a true "high-performing" school and throughout the remainder of this report we analyse the data both with and

without the Baropodi scores to show that when it is removed from the analysis, the data for other schools generally shows an even better "fit" with the expected distribution (i.e. around the 50th percentile).

Figure 10

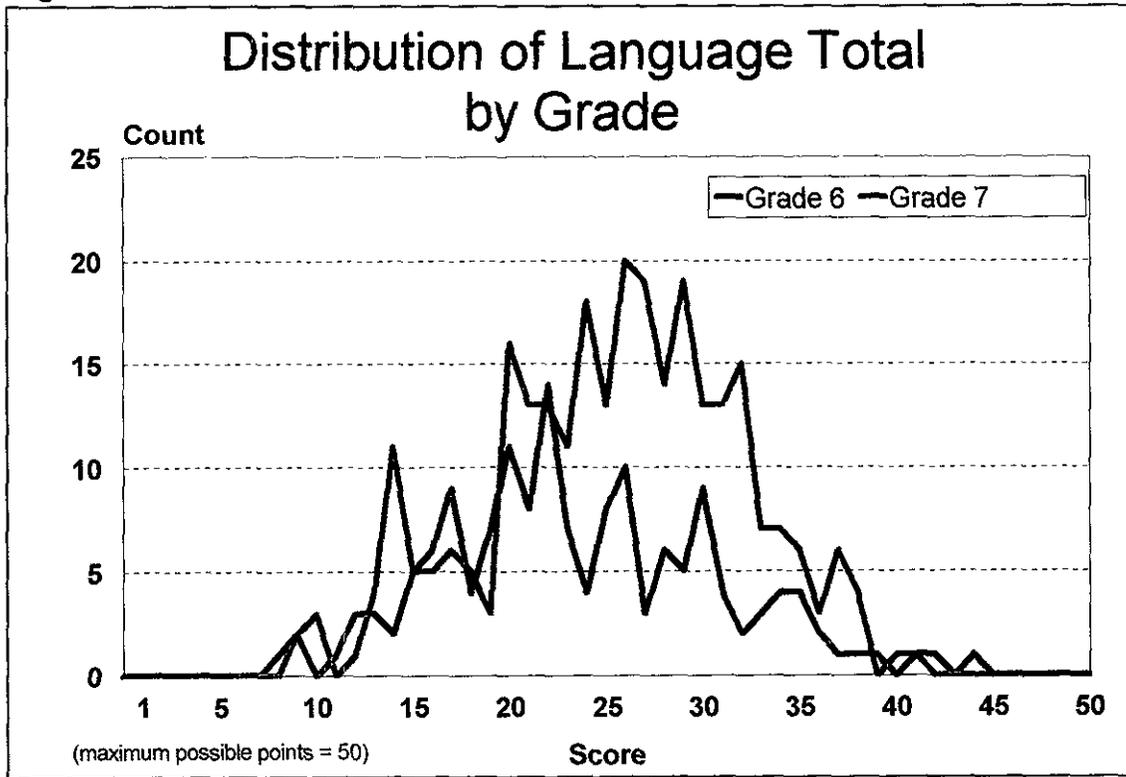


Figure 11

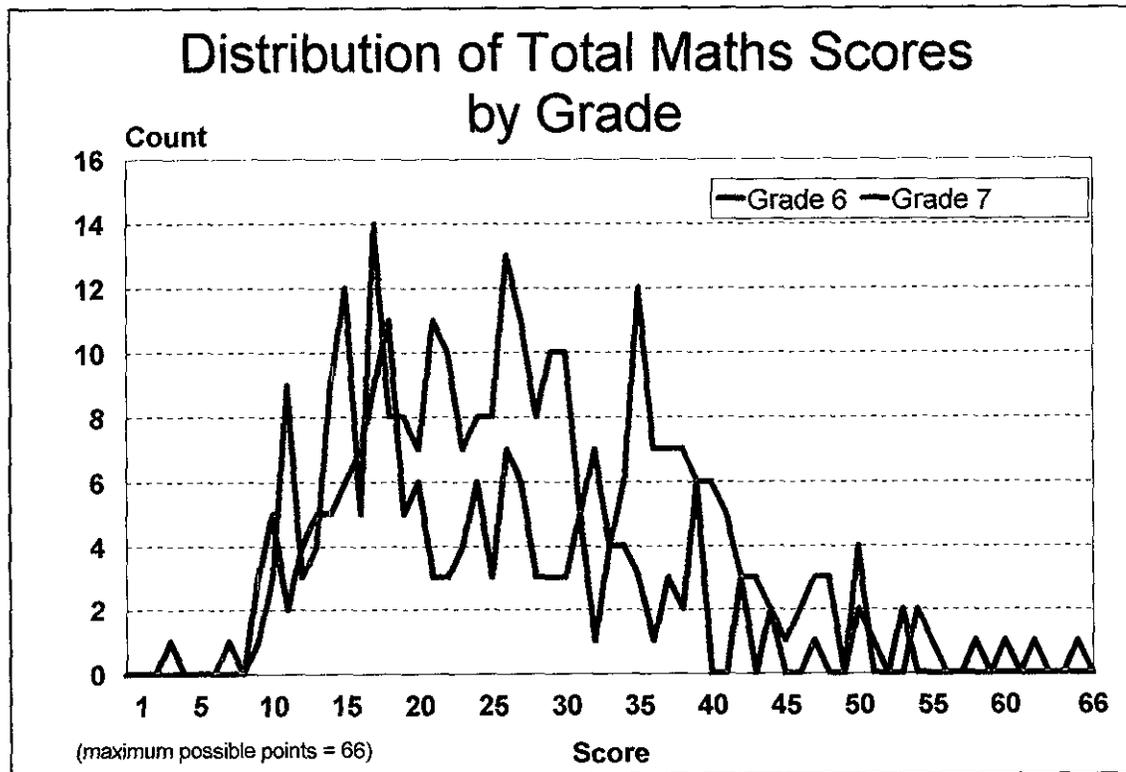


Figure 12

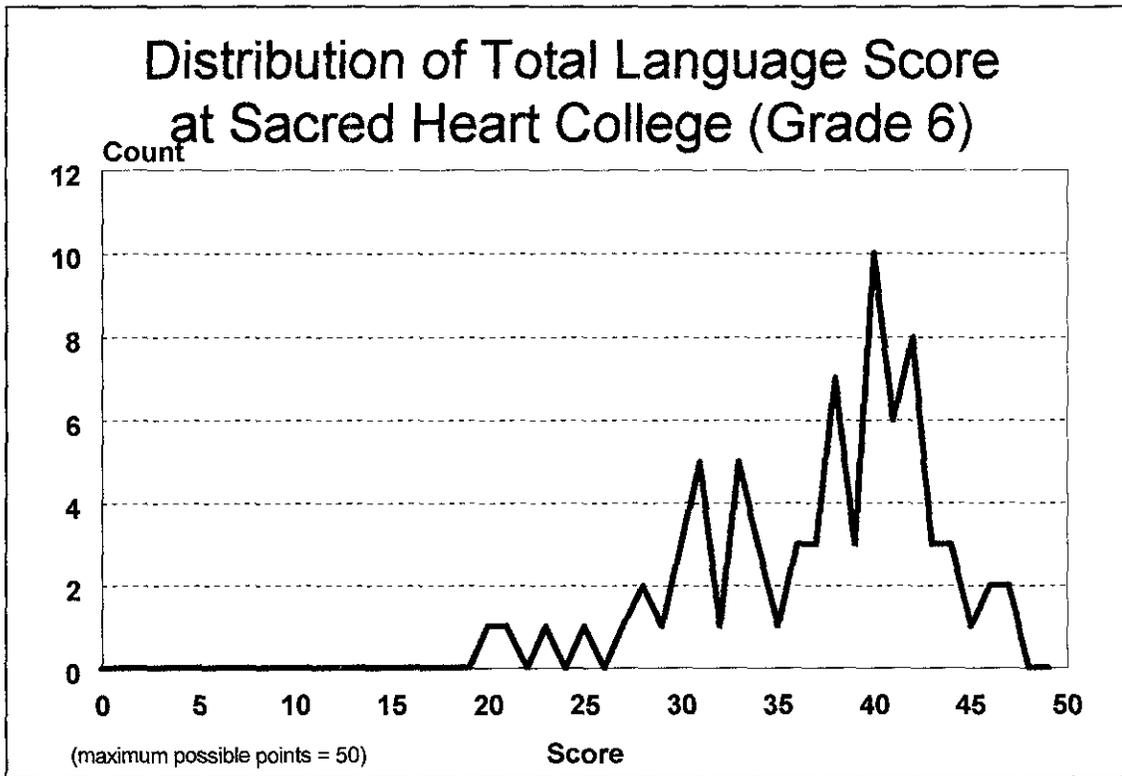
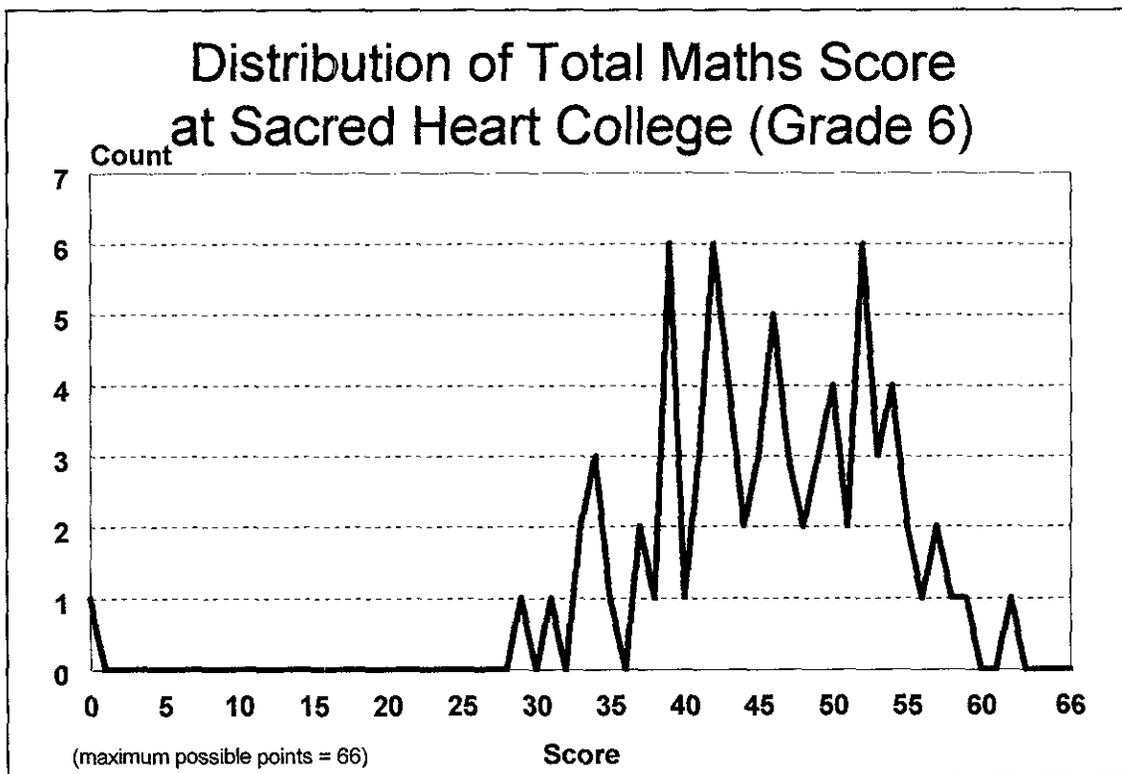


Figure 13



pupil's ability. The point was made, however, that the prompts provided did not sufficiently stimulate or challenge the best pupils in each class. Moreover, SHC teachers believed that the writing samples produced by these high-performing pupils did not adequately reflect their abilities.

A constraint which can affect reliability was the fact that the three teachers involved had administered the tests differently. This discrepancy points to a need to brief teachers on the administration of the tests, in order to assure standardisation of test administration.

The Maths test was administered to the same group of Grade 6 learners at SHC. Again, the distribution of scores shows that the Grade 6 learners outperformed all the other Grade 7 learners at the high performing schools, although the test was a bit more challenging as evidenced by the clustering of the scores at the 65th-70th percentile ($p=.000$) (see Figure 13).

Conclusions regarding Test Validity: the Language Total Scores and the Maths Total scores are statistically significant normal distributions (See Appendix E) indicating that they are valid in discriminating between poor, average, and good achievement.

With respect to the relative ease and difficulty of the tests, when Baropodi is removed, neither test is clustered around the expected mean, indicating that they are still a bit too hard or easy than would be expected. In the case of the English Language test the distribution of total scores indicates that the test is easier than desired ($p=.000$) and in the case of the Maths test the distribution of total scores indicates that it is slightly harder than desired ($p=.014$).

However, as is discussed in the following sections, three of the four components of the Maths test are actually too easy and the total score is being "pulled down" by one very difficult section (Maths 3: Spatial Perception and Reasoning). These findings confirm the SHC expert opinion that the English Language test is not challenging enough to discriminate at progressively higher levels of achievement. Therefore, the English Language test needs to be revised by including more difficult items.

SHC expert opinion on the validity of the Maths test is not consistent with the results achieved on Maths total scores, but as will be seen in the following discussion on the individual Maths components, three of four components are actually too easy, thereby confirming SHC expert opinion.

5.2.3 Effects of Home Language on English Language Test Performance

It was not possible to undertake statistical analysis of the effects of home language on learners' performance of the English Language Test because the vast majority of learners spoke Pedi or Venda and there were insufficient numbers of other-language speakers to allow for meaningful comparisons between the different home languages (see Table 1). If additional funding is granted in the future, the English Language Test will be administered to greater numbers of other language speakers to ascertain the effects of home language on performance on the English Language test.

Table 1. Learners' Home Language (Grade 7 only)

Language	N =	Percent
Pedi	186	66.91
Venda	41	14.75
English	28	10.07
Zulu	8	2.88
Tsonga	3	1.08
Tswana	3	1.08
Ndebele	2	0.72
S. Sotho	2	0.72
Siswati	2	0.72
Xhosa	2	0.72
Afrikaans	1	0.36
Total	278	100.00

5.2.4 Performance of the Individual Elements of the Language Test

Language 1: Lang I measures reading comprehension skills. Figure 14 through Figure 16 below show the distribution of scores from 0 to the full possible of 12. The statistically significant normality of the distribution (see Appendix E) indicates that Lang I generally discriminates between a wide range of abilities in reading comprehension⁷. Unfortunately, the observed mean is significantly higher than the expected mean (and the most frequently found scores were in the 6 to 9 range -- i.e. higher than expected), indicating that the Lang I section may need to be modified by the inclusion of a number of more challenging items in order to reflect a more "normal" distribution around the 50th percentile. This point is underscored by the fact that a higher than expected percentage of learners in the sample was able to achieve full scores on this section of the test.

Analysis of performance by school (Figure 16) indicates that the scores achieved by Baropodi Primary approximate a normal distribution, with a median score of 5 out of a possible 12. Given that this school was found to be the poorest performing school in the sample, this again highlights the finding that this section of the Language Test was not sufficiently challenging. A greater number of "second" and "third" order comprehension questions needs to be included in this section of the test. In contrast, the median score of 10 for St. Theresa's further emphasises this need, as the test was clearly not sufficiently challenging for this group of learners. Of course, the fact that St. Theresa's has a larger proportion of English First Language learners than the other sample schools, may possibly account for their better performance on this component.

Learner attainment at Little Bedfordview Primary demonstrates a greater degree of variation than is evident at the other schools (Figure 16). This could have two possible explanations. Either the school has a wider than normal range of pupil ability, or different English teachers are responsible for different levels of learner achievement. The role of the individual teacher in affecting learner achievement requires further investigation.

Ranking of schools by median score indicates that St. Theresa's was the highest performing school on this component of the Language Test, followed sequentially by Little Bedfordview, St. Scholastica's and Baropodi.

⁷ That some pupils got very low scores on the reading comprehension component, suggests that they have little or no skills in reading comprehension in English, even close to the end of their primary school years, which is worrying indeed.

Language 2: Lang2 measures learners' skills in interpreting and responding to information presented in a visual mode. This section assesses also learners' ability to respond effectively to "real-life" problems.

Figure 17 through Figure 19 show the distribution of scores for Lang2. Here again, a wide range of ability is seen. Although the low Baropodi scores significantly pull down the overall mean, once this school is removed from the data set, this section of the test does appear to present adequate challenge ($p=.39$). Furthermore, no pupils achieved full scores, although few achieved scores at the top end of the range. This is the only section of the English Language test which does not require any revision to make it more challenging.

Ranking of schools by median score again places St. Theresa's in the top position, followed by Little Bedfordview. St. Scholastica's performance on this component is relatively worse than performance on Lang1 (reading comprehension), suggesting that reading comprehension is an aspect of instruction neglected at St. Scholastica's.

Language 3: Lang3 measures skill in writing.

Figure 20 through Figure 22 below show the distribution of scores for Language 3. As with the other components of the Language Test, a wide range of skill is evident. Although the test scores are clustered around the expected mean, once the Baropodi scores are removed from the analysis (Figure 21), test scores assume a significant deviation above the expected mean ($p=.000$). This again suggests that this section of the Language Test does not pose sufficient enough challenges for high performing learners.

Ranking of performance by school indicates that, relative to other areas, writing is an area of weakness for St. Theresa's and an area of strength for Little Bedfordview. These findings suggest that writing is an aspect of instruction which has been neglected at St. Theresa's and privileged at Little Bedfordview. This information can be used to redirect or modify instructional programmes at each school.

General Conclusions: Appendix E shows that all the distributions of scores for the components of the English Language test are significantly normally distributed, thereby indicating that the components generally discriminate between poor, average, and high achievement. However, Table 2 below, summarising the t-test statistics, shows that after Baropodi Primary is removed from the analysis, two of three components (along with the total score) are significantly "easier" than would be expected, indicating that the test generally requires more challenging items to better discern high-performing ESL learners from others. Language 1 (reading comprehension) was easy even for Baropodi learners, while Language 2 (visual literacy) and Language 3 (writing) show an increase in overall learner achievement once the Baropodi results are removed from the analysis (from "too-hard" to neutral in the case of Language 2 and from neutral to "too easy" in the case of

Language 3).

Figure 14

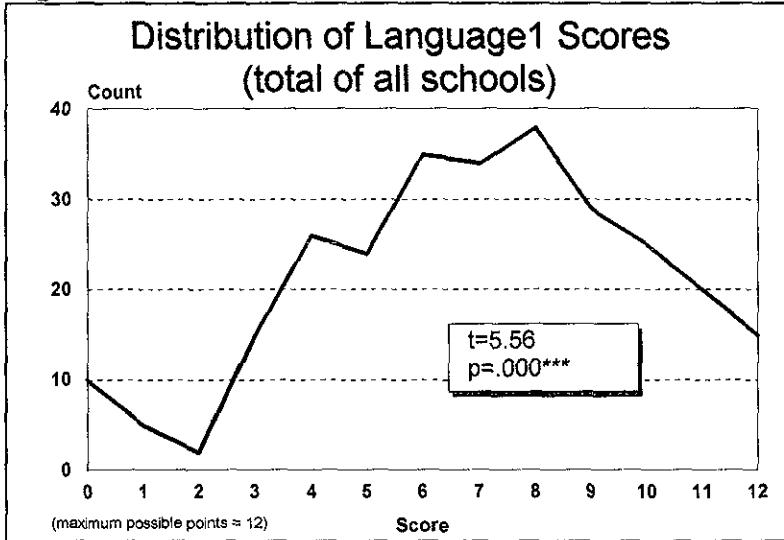


Figure 16

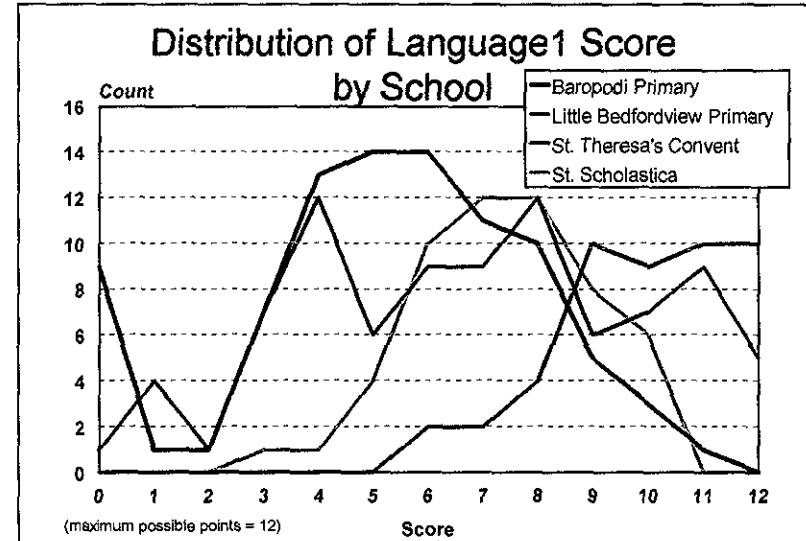
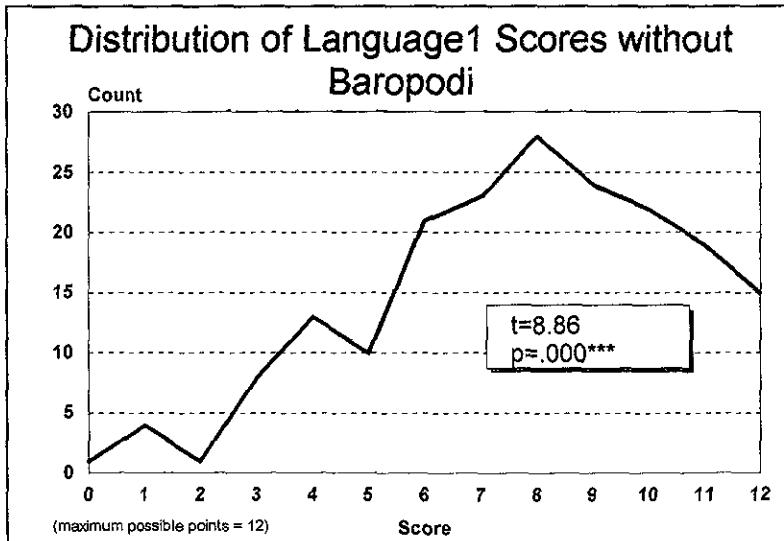


Figure 15



Language 1 – Reading Comprehension

School Name	Median Score	Rank
Baropodi Primary School	5.0	4
Little Bedfordview	7.0	2
St. Theresa's Convent	10.0	1
St. Scholastica's Primary School	6.0	3

Figure 17

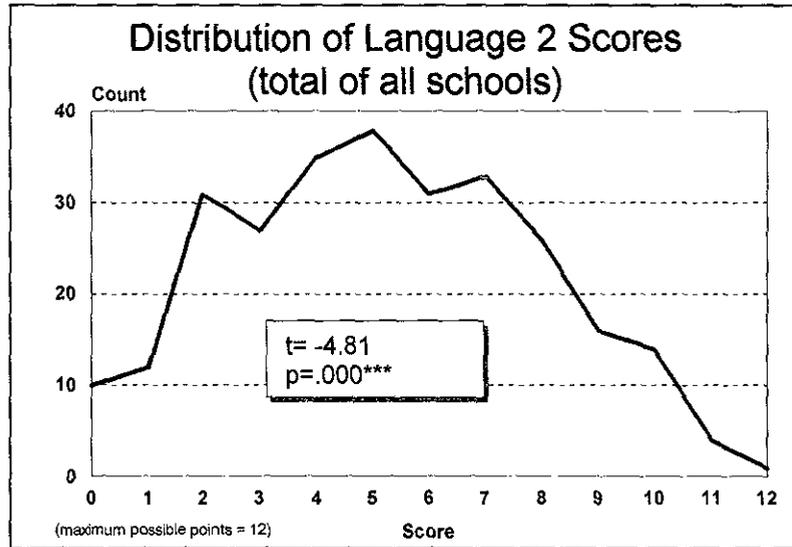


Figure 19

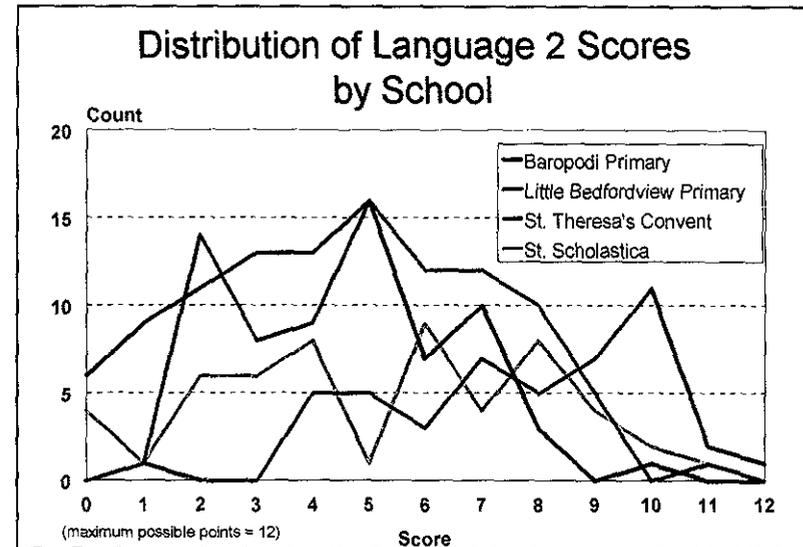
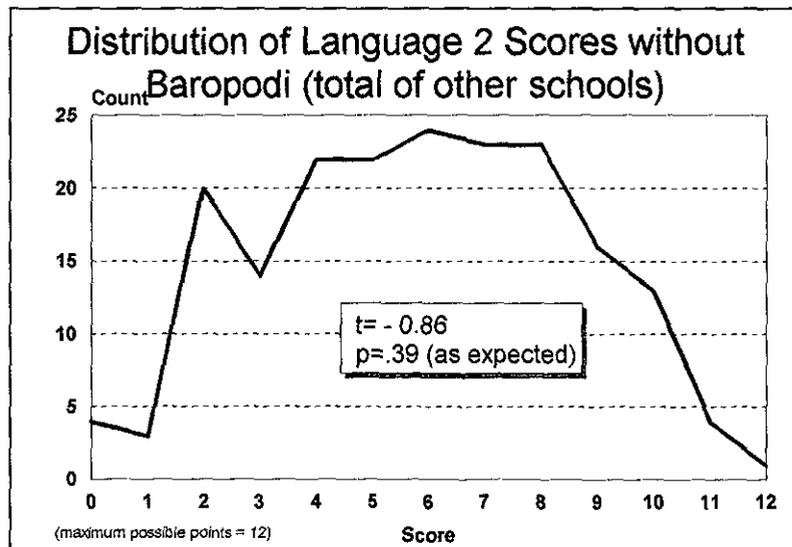


Figure 18



Language 2 – Visual Literacy

School Name	Median Score	Rank
Baropodi Primary School	4.0	3
Little Bedfordview	5.5	2
St. Theresa's Convent	8.0	1
St. Scholastica's Primary School	4.0	3

Figure 20

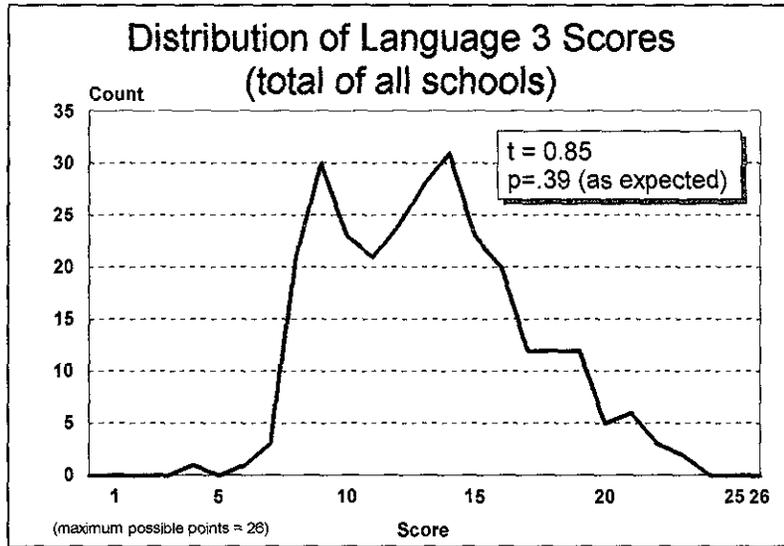


Figure 22

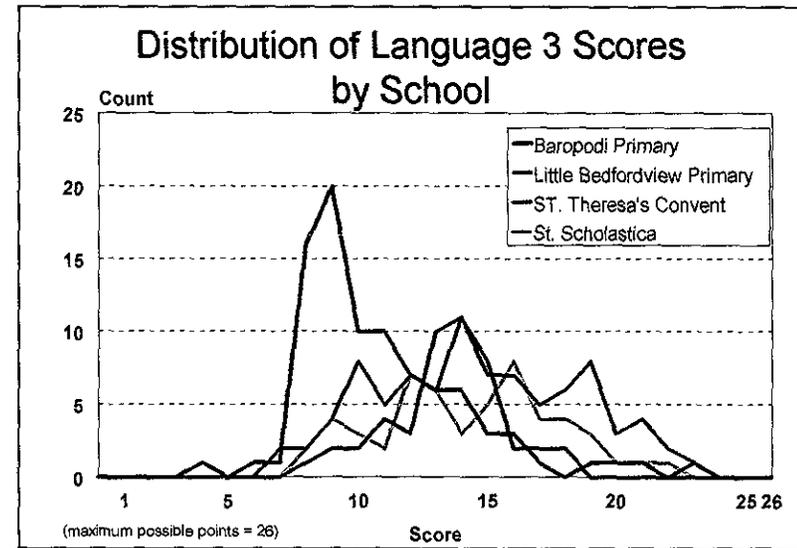
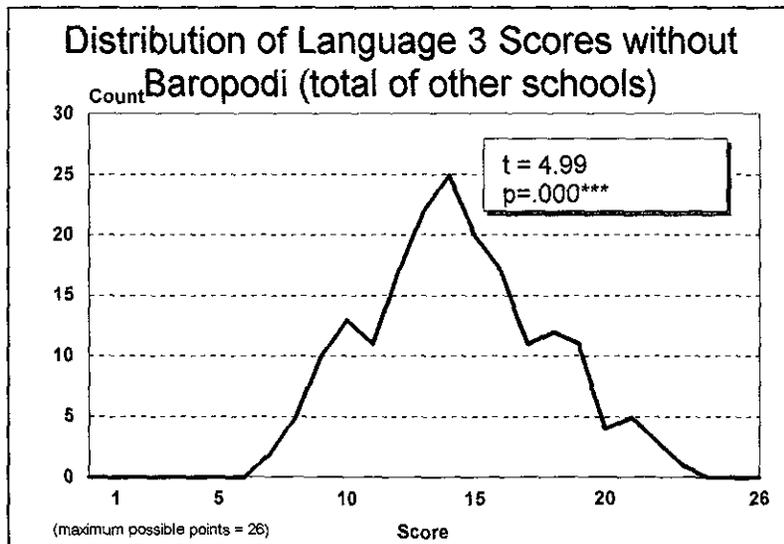


Figure 21



Language 3 – Writing

School Name	Median Score	Rank
Baropodi Primary School	10.0	3
Little Bedfordview	14.0	1
St. Theresa's Convent	14.0	1
St. Scholastica's Primary School	13.0	2

Table 2. Summary Statistics for English Language Test components

	Expected Mean	1-sample t-test on all Schools (N=278)			1-sample t-test on schools except Baropodi Primary (N=198)		
		Actual mean	t=	p=	Actual mean	t=	p=
Language 1	6.0	6.98	5.56	.000	7.76	8.86	.000
Language 2	6.0	5.21	-4.81	.000	5.83	-0.86	.39
Language 3	13.0	13.91	0.85	0.39	14.24	4.99	.000
Language Total	25.0	25.38	0.98	0.33	27.84	7.48	.000

Note: *bolded cells indicate statistically significant differences (either higher or lower as indicated by the "t" results) from the expected mean.*

Despite the fact that a significant difference was found between the scores of Grade 6 and Grade 7 learners at the sample schools, the above data underscores the opinions of Ms. De Gouveia at SHC that the English Language test does not adequately discriminate between poorer and higher performing learners.

Therefore it can be concluded that while the scores are normally distributed, the observed means are generally above the expected means (the 50th percentile score for each component) confirming that overall the test is not challenging enough to be used as a standard test and should be revised accordingly.

Nevertheless given that the individual components of the English test have been constructed along the lines of Curriculum 2005 objectives, the individual components of the tests have the potential for providing valuable diagnostic information regarding learner attainment in each component area.

5.2.5 Performance of the Individual Elements of the Maths Test

Maths 1: Maths1 measures learners' ability to perform the four basic operations of Mathematics at a level appropriate to Grade 7.

Figure 23 through Figure 25 show the distribution of scores for Maths1. Scores indicate a wide range of skill in this area, suggesting that it is not correct to assume that, by this stage in the school programme, most learners have mastered the basic computational skills in Maths. Despite this, the distribution of scores and the observed mean was as expected for all schools, yet once Baropodi was removed, the observed mean was significantly higher than expected ($p=.000$), suggesting that this component is in fact "too easy".

Achievement levels at Baropodi Primary are particularly poor, as indicated by a median score of 3 out of a possible total of 12. The bi-modal distribution at Little Bedfordview indicates that some Grade 7 pupils have an adequate level of mastery of basic computation skills, while a substantial number of pupils are not yet at this level. Ranking of schools by median score indicates that, as with the Language Test, learners from St. Theresa's were the most proficient, followed by Little Bedfordview.

Maths 2: Maths2 measures learners' ability to use the language of mathematics and understanding of basic mathematical concepts.

Figure 26 though Figure 28 below show the distribution of scores for Maths2. In line with other components of the tests, performance of the Baropodi pupils was very poor, as indicated by the median score of 2 out of a possible 4. Baropodi's poor performance contributed to no significant difference in the observed mean compared to the expected mean,

but once it was removed from the analysis, the mean of 5 (out of a possible 9) represents a significant deviation above the expected mean of 4.5 ($p=.0009$), showing again that this component is "too easy" and should be revised to make it more challenging by adding more higher-level items.

Performance of St. Scholastica's learners was little better than Baropodi, with a median score of 3. In contrast performance of the St. Theresa's pupils was good, as indicated by the median score of 7, and by the fact that a number of St. Theresa's pupils were able to achieve full score on this component. The performance of the Little Bedfordview pupils was variable, with scores spread across the range from 0 to the full score.

Maths 3: Maths3 measures learners' skills in spatial perception and reasoning. This component has some reliance on syllabus and content covered in the classroom.

Figure 29 through Figure 31 show the distribution of scores for Maths3. All students at all schools performed poorly on this component. The clustering of scores around the 20th percentile of the test ($p=.000$) indicates that learning in this area is weak to non-existent. In addition, the time of year when the test is written can be expected to affect results - the later in the year, the greater the acquisition of skills. If, as was the case here, testing is conducted fairly early in the year, it would be expected that performance would be poor - indeed this is what was found. In addition, the lack of apparatus for the teaching of spatial perceptual skills at most disadvantaged schools will contribute to poor performance in this area.

There was no change in the statistical significance of the deviation of the observed mean from the expected mean when Baropodi was removed (although the observed mean did increase from 4.5 to 5.1).

Maths 4: Maths 4 measures learners' skills in problem-solving and deductive thinking through the mode of mathematics.

Figure 32 through Figure 34 show the distribution of scores for Maths 4. The scores on this component are normally distributed, confirming that this component has the power to distinguish between different levels of attainment. However, once Baropodi was removed, the mean was significantly higher than expected ($p=.000$) suggesting again that this component is "too easy" and requires the introduction of more challenging items.

Once again, performance at Baropodi was poorest as indicated by the median score of 9 out of a possible 25. Ranking of schools by median score produces the same picture, with St. Theresa's in top position with a median score of 17.

Performance of Learners without Maths 3: Given that learner performance on Maths 3 was poor in all schools, the distribution of total Maths scores without Maths 3 was analysed (see Figure 35 through Figure 37). When Maths 3 is removed, the distribution of Maths scores has a better fit around the expected mean (the 50th percentile) than when Maths 3 is part of the total score (see Figure 5 on page 13). And when Baropodi Primary is removed from the analysis, the scores cluster significantly above the expected mean ($p=.000$). This further confirms that the other Maths test components are not challenging enough. It also suggests that the types of questions in Maths 3 have not been taught at this point in the school year at any school in the sample.

Figure 23

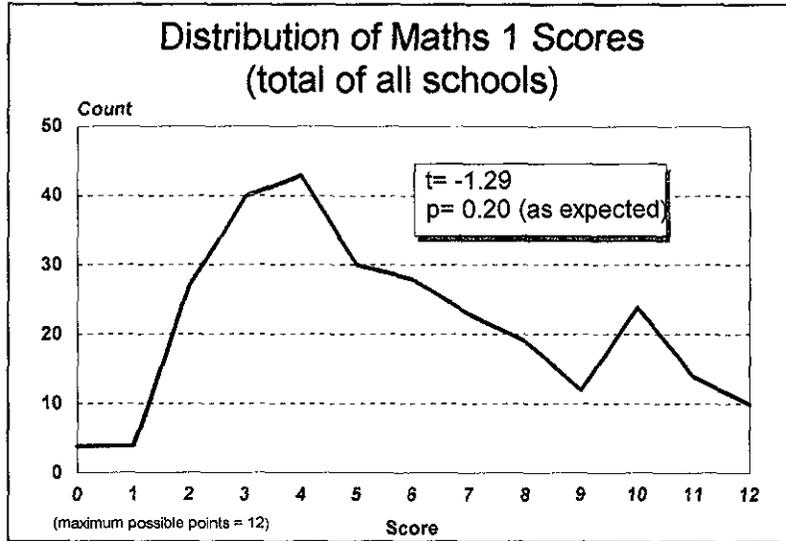


Figure 25

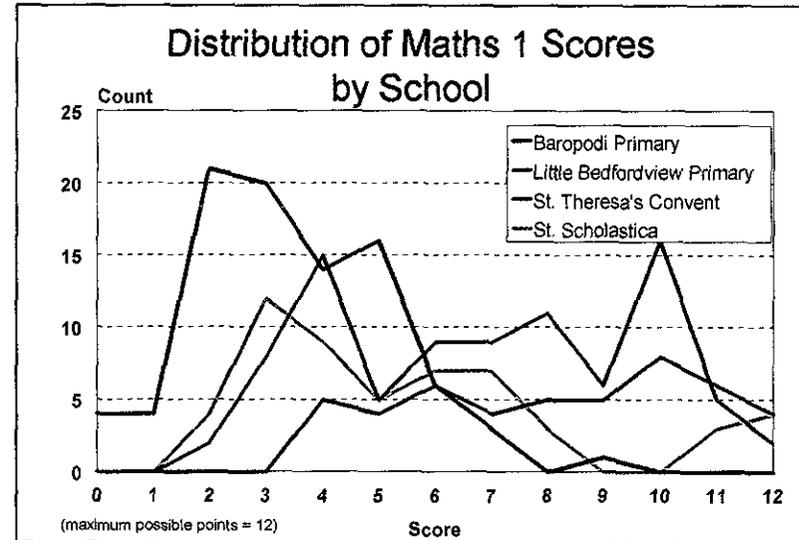
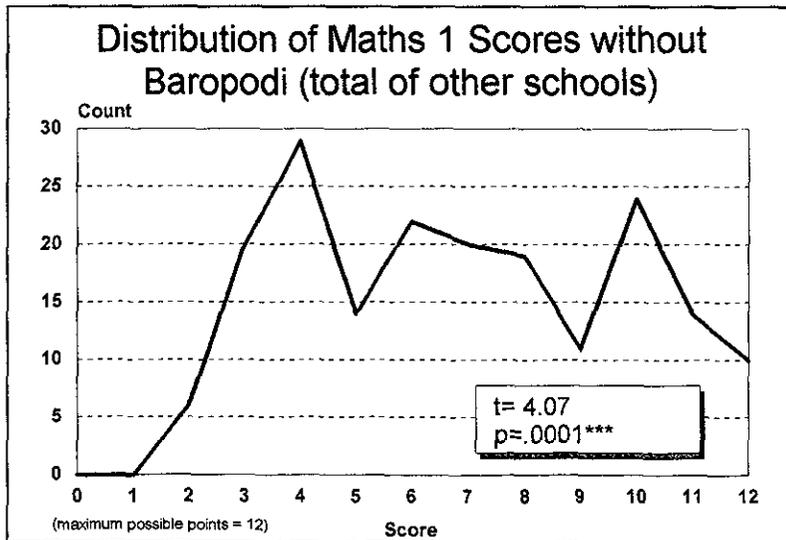


Figure 24



Maths 1 – Basic Operations

School Name	Median Score	Rank
Baropodi Primary School	3.0	4
Little Bedfordview	6.0	2
St. Theresa's Convent	8.0	1
St. Scholastica's Primary School	4.0	3

Figure 26

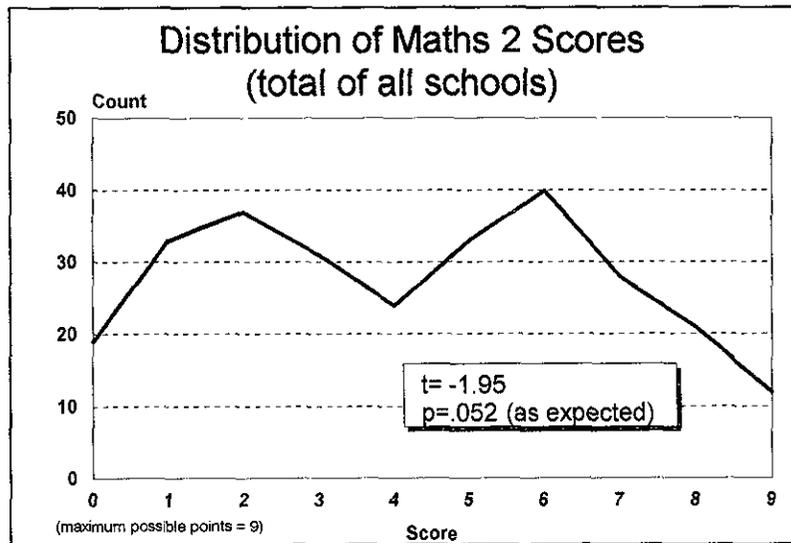


Figure 28

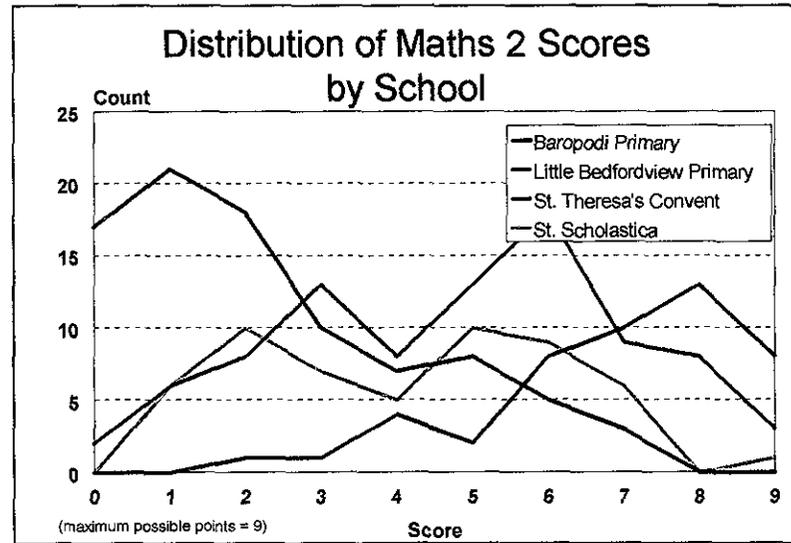
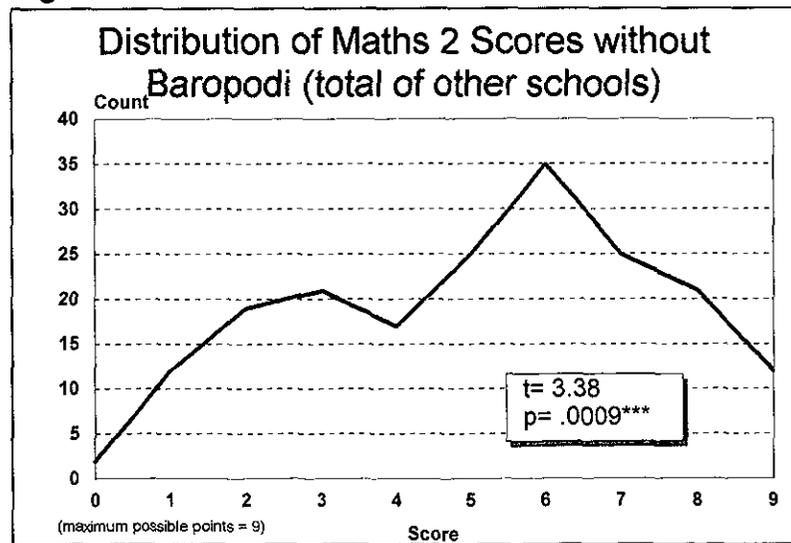


Figure 27



Maths 2 – Language of Mathematics

School Name	Median Score	Rank
Baropodi Primary School	2.0	4
Little Bedfordview	5.0	2
St. Theresa's Convent	7.0	1
St. Scholastica's Primary School	3.0	3

Figure 29

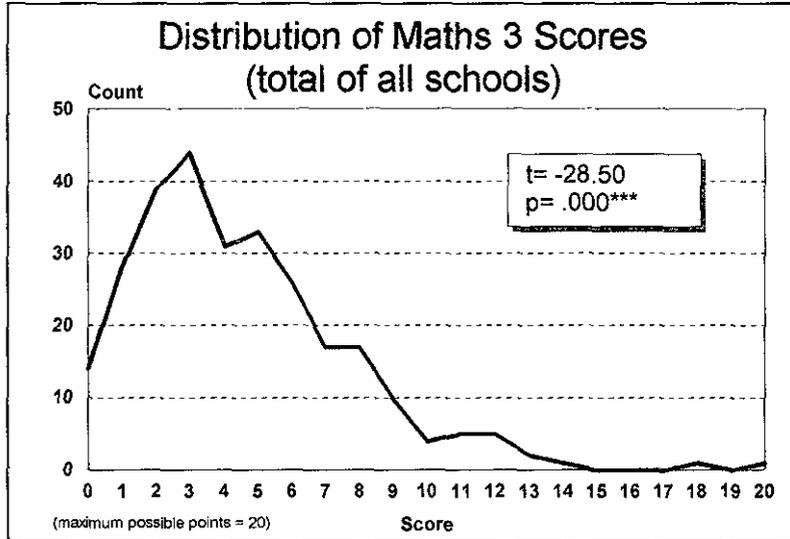


Figure 31

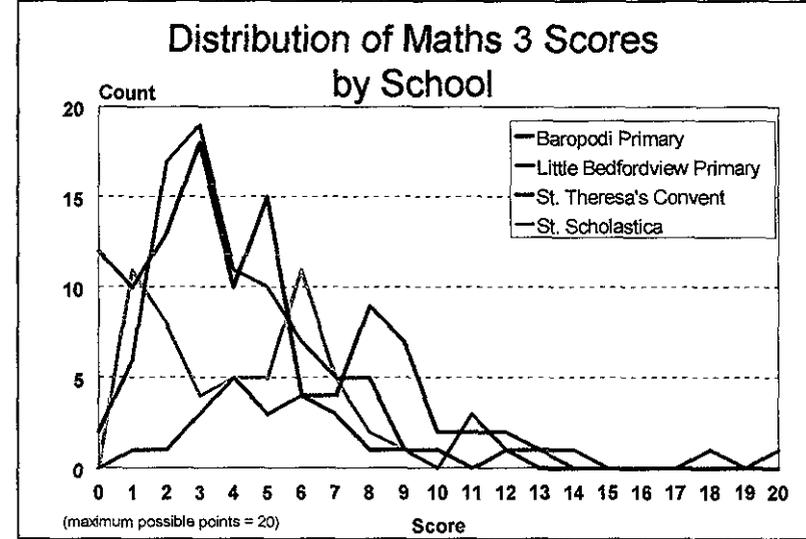
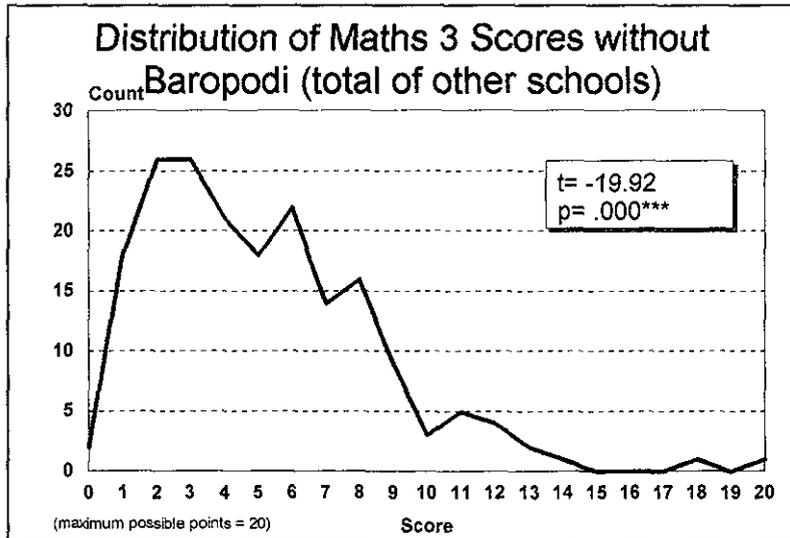


Figure 30



Maths 3 – Spatial Perception and Reasoning

School Name	Median Score	Rank
Baropodi Primary School	3.0	3
Little Bedfordview	4.0	2
St. Theresa's Convent	7.5	1
St. Scholastica's Primary School	3.0	3

Figure 32

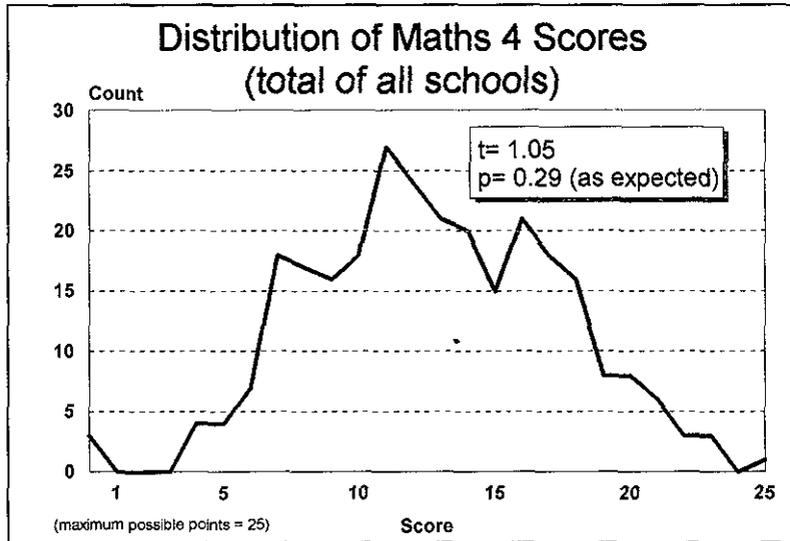


Figure 34

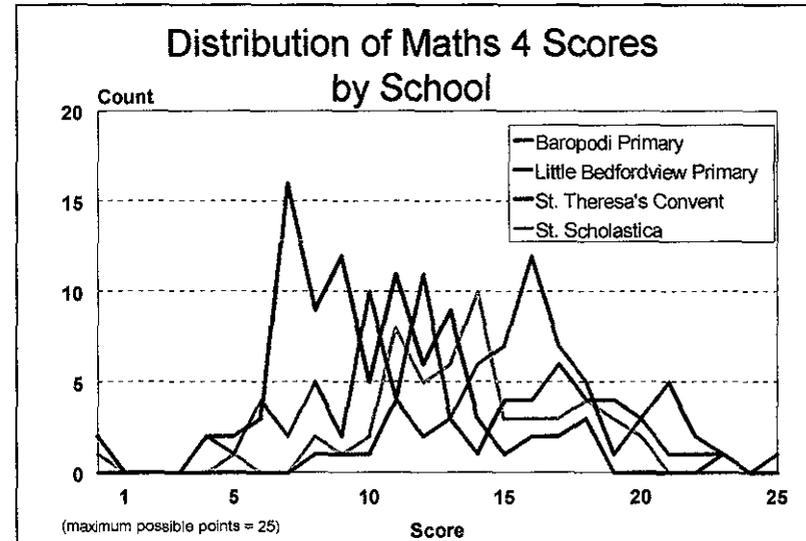
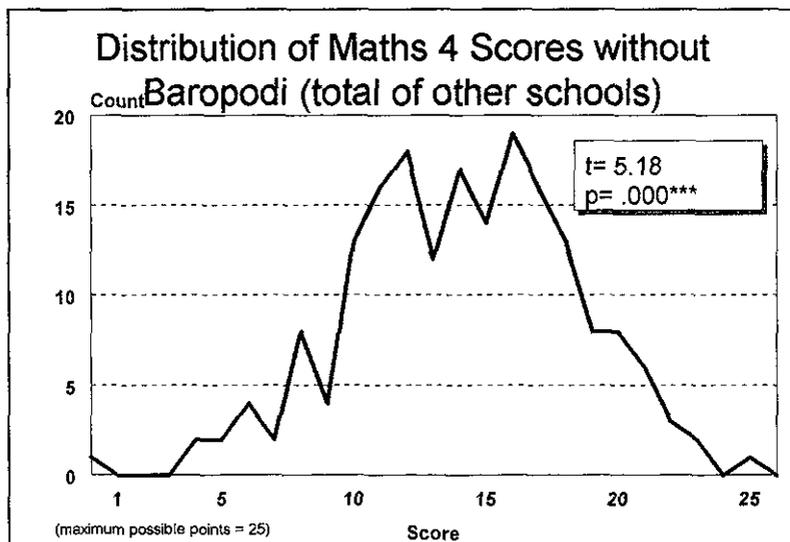


Figure 33



Maths 4 – Problem solving and deductive thinking

School Name	Median Score	Rank
Baropodi Primary School	9.0	4
Little Bedfordview	14.0	2
St. Theresa's Convent	17.0	1
St. Scholastica's Primary School	12.0	3

Figure 35

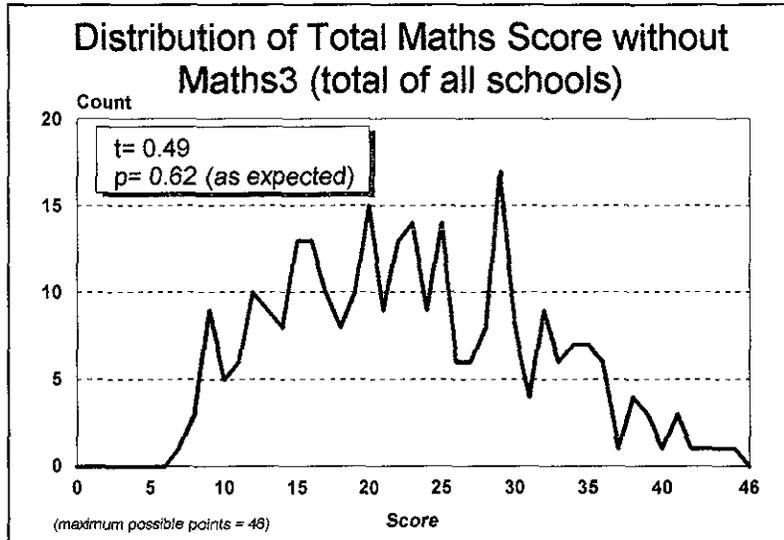


Figure 37

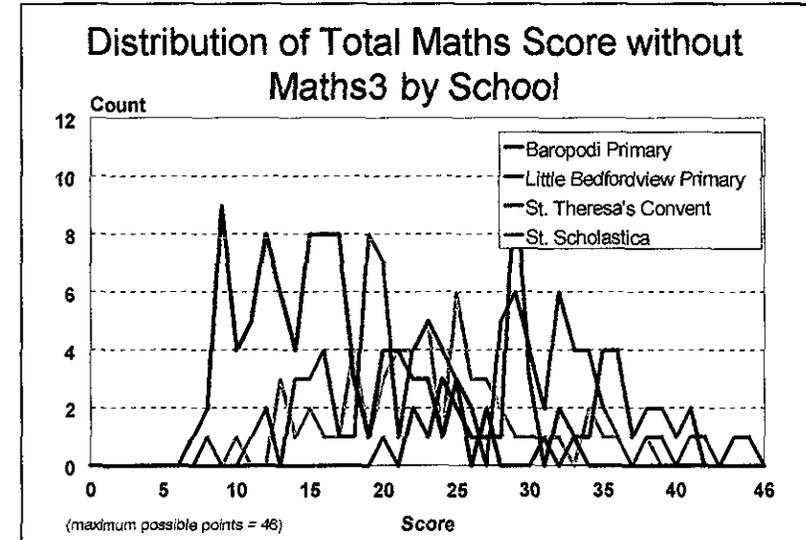
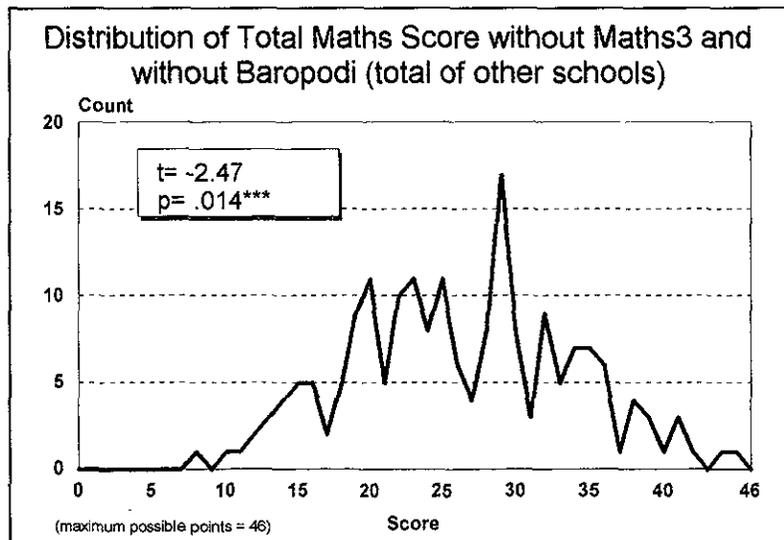


Figure 36



Maths without Maths 3

School Name	Median Score	Rank
Baropodi Primary School	9.0	4
Little Bedfordview	14.0	2
St. Theresa's Convent	17.0	1
St. Scholastica's Primary School	12.0	3

St. Bamabas feels strongly that spatial perceptual understandings and skills are an important part of Maths and that this component of the test should be retained as an indicator of what should be taught at schools, even though this might not be the case at present, and even though this might lead to lower performance on the Maths test overall.

General Conclusions: Appendix E shows that (with the exception of Maths 1), the distributions of scores for the components of the Maths Test are normally distributed - thereby indicating that the components generally discriminate between poor, average, and high achievement.

However, Table 3 below, summarising the t-test statistics, shows that after Baropodi Primary is removed from the analysis, the mean for every component (except for Maths 3) is significantly above the expected mean. This is true even when Maths 3 is removed from the computation of the total score, indicating that overall the test is "too easy" and generally requires more challenging items to better discern high-performing ESL learners from others.

Despite the fact that a significant difference was found between Grade 6 and Grade 7 learners at the sample schools, these data again underscore the opinions of Ms. De Gouveia at SHC that the Maths test also does not adequately discriminate between poorer and higher performing learners.

Table 3. Summary Statistics for Maths Test components

	Expected Mean	1-sample t-test on all Schools (N=278)			1-sample t-test on schools except Baropodi Primary (N=198)		
		Actual mean	t=	p=	Actual mean	t=	p=
Maths 1	6.0	5.76	-1.29	.200	6.85	4.07	.0001
Maths 2	4.5	4.19	-1.95	.052	5.08	3.38	.0009
Maths 3	10.0	4.53	-28.50	.000	5.14	-19.92	.0000
Maths 4	12.5	12.78	1.05	.290	14.11	5.18	.0000
Maths Total	33.0	27.28	-8.84	.000	31.18	-2.47	.0140
Maths Total without Maths 3	23.0	22.75	-0.49	.062	26.04	5.53	.0000

Note: bolded cells indicate that statistically significant differences (either higher or lower as indicated by the "t" results) from the expected mean.

Therefore, while the scores are normally distributed, the observed means are generally above the expected means (the 50th percentile score for each component) confirming that the overall Maths test is not challenging enough to be used as a standard test and should be revised accordingly.

Nevertheless given that the individual components of the Maths test have been constructed along the lines of Curriculum 2005 objectives, the components have the potential for providing valuable diagnostic information regarding learner attainment in each component area for each school.

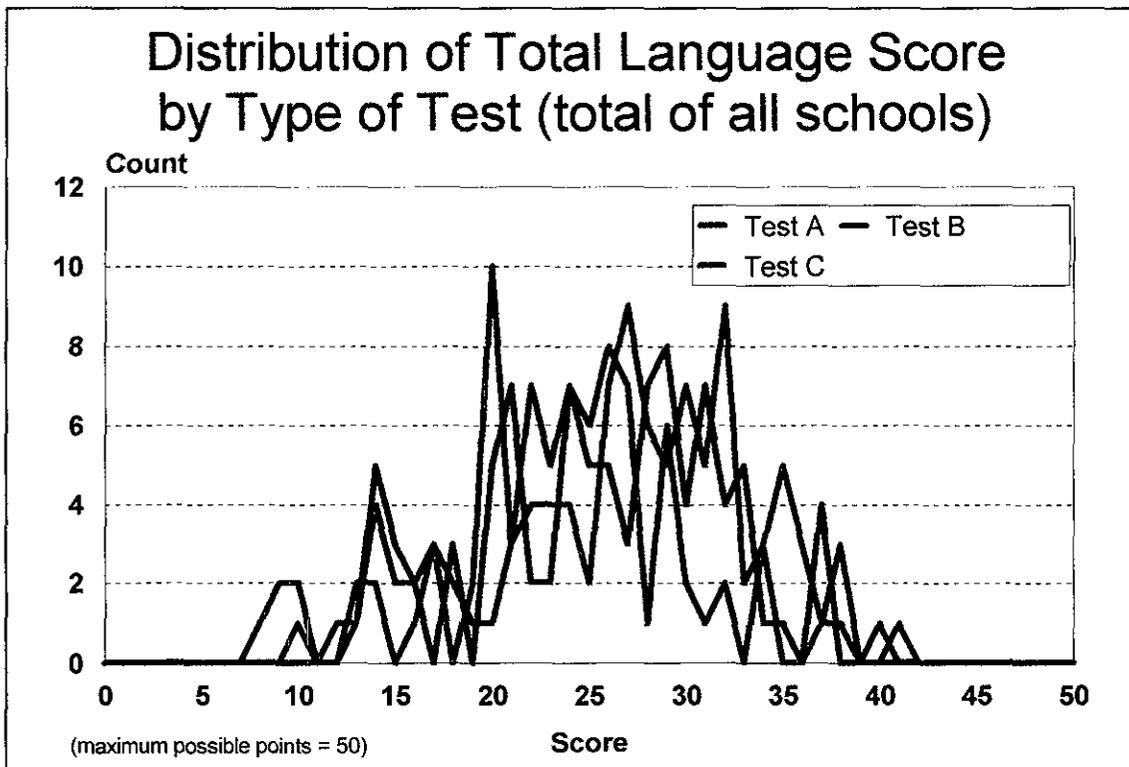
5.2.6 Equivalency of the Three Test English Language Forms (Tests A, B, and C)

As mentioned in the methodology section, three forms of the English Language test were designed (Test A, B, and C) for the purposes of allowing for testing and re-testing with alternative forms of the same test. Figure 38 below presents the distribution of total

alternative forms of the same test. Figure 38 below presents the distribution of total English Language scores by type of test. The graph shows that Test B is generally more "difficult" than Test A or Test C (which is the "easiest" test form). However, these differences are not significantly different ($p=.08$) suggesting that the different forms of the test are indeed comparable.

Given the relative comparability of the three tests, one form could be used at the beginning of the school year and another during or at the end of the school year to measure improvements or changes in learner performance without the risk of interference from a 'test-wise' factor. Tests A or C should be used at the beginning of the year and Test B at the end of the year.

Figure 38



Maths Test

Finally, comparisons between the performance of girls and boys was examined to see if the test had inherent gender biases. Figure 39 through Figure 41 below show no significant gender difference in performance on the Language Test ($p=.26$), Maths total ($p=.31$) or Maths total without Maths3 ($p=.31$). This indicates that the tests do not discriminate on the basis of gender. These findings are consistent with findings from the TIMSS study that showed that in South Africa, unlike other countries, there were no significant differences between the performance of boys and girls in Maths and Science.

Figure 39

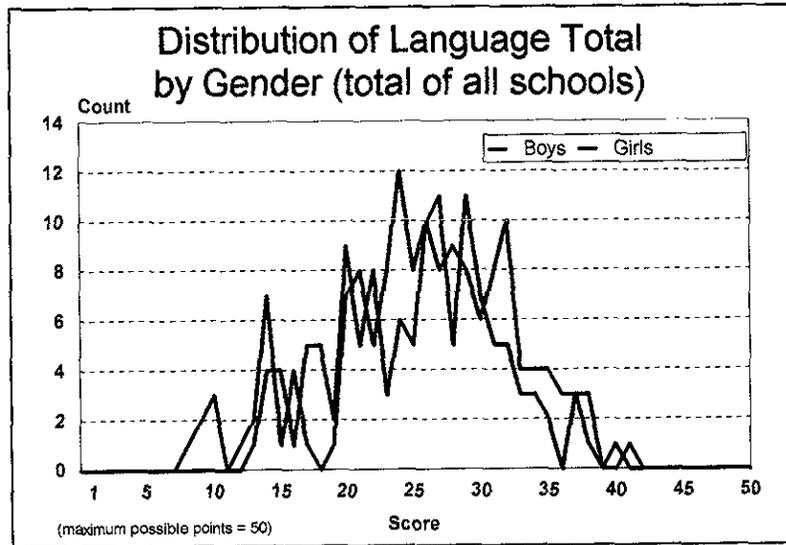


Figure 41

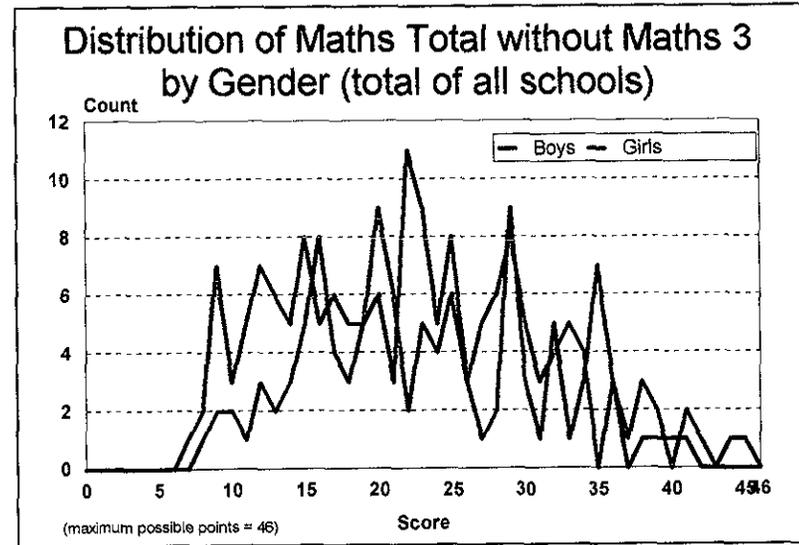
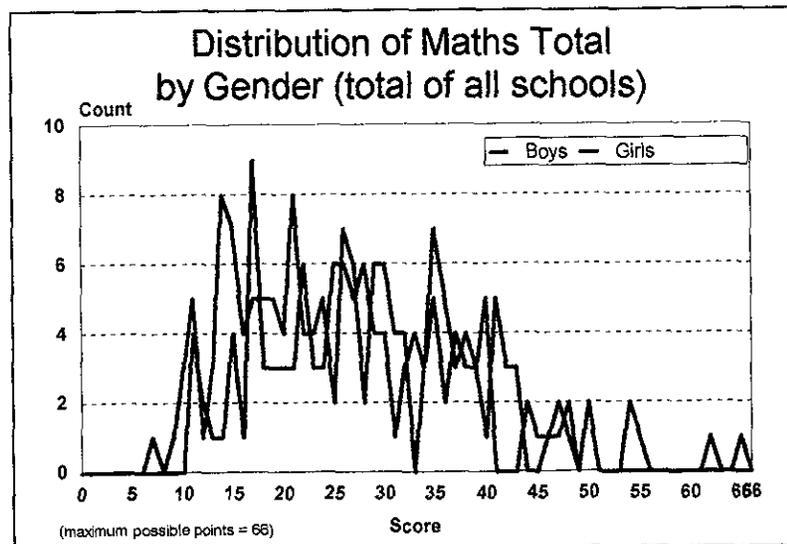


Figure 40



5.3 Examination of Tests in Current Use at Sample Schools

5.3.1 English Tests in Current Use

The emphasis of these test (see Appendix C) is on assessing splinter skills (adjectives, verbs, prepositions, tenses, indirect speech, negative form, passive voice, pronouns, opposites, etc) rather than the assessment of communicative competence. C2005 emphasises a more holistic approach to receptive and expressive language rather than the mastery of skills in isolation. Nevertheless, the level of assessment of these splinter skills is largely appropriate at the Grade 7 level for ESL learners.

With regard to reading comprehension, all questions are first level - that is literal comprehension -- and there are no higher level questions that measure inferential or deductive thinking skills. The test do not include any further kinds of literacy, such as visual literacy, which are part of the St. Barnabas tests developed under this study.

5.3.2 Maths Tests in Current Use

The emphasis on these tests (see Appendix C) is on basic calculation, although at an appropriate Grade 7 level. Most of the items, however, require first-level responses (e.g. name, identify, recall, reproduce) and do not incorporate higher level thinking skills which would test conceptual understanding of mathematics.

Although there are items which measure competence in calculating area and perimeter, when these items are adapted to a visual modality in the Maths 3 section of the St. Barnabas tests (i.e asking for the calculation of the "area of the fence around the school") the learners were unable to respond. This suggests that learners are able to respond in a rote fashion, but display little understanding of the concepts.

6. DISCUSSION

A number of important findings have emerged from this research project:

1. Scores on most components of the St. Barnabas tests are normally distributed. This finding indicates that the tests generally provide a valid indication of the range of abilities for disadvantaged ESL learners at a Grade 7 level. 6.
2. Critical comment from an expert in the field of assessment was sought (see Appendix F: Comment from Mr. Nick Taylor of Joint Education Trust). Mr. Taylor disputed the notion that a test based on a sample of four schools could be regarded as standardised. His recommendation was that a more valid claim would be that the tests are norm referenced with respect to three high achieving ESL schools in the Northern province and Gauteng. Further pilot testing of the instruments in other schools in other provinces in South Africa will lead to more valid standardisation.
3. Because the details of the Specific Outcomes for Maths and English in the Intermediate Phase were not completed at the time of constructing these tests, criterion-referencing was limited. However, future "item maps" linking each test item to the Specific Outcomes will further validate the content of the tests.

4. Unfortunately, the St. Barnabas tests are generally not challenging enough even for disadvantaged ESL learners in South Africa. In order for these tests to be accepted as standard assessment instruments, they would need to be revised by the inclusion of more challenging items which would bring the means of the distribution more in line with the expected means (at the middle range of the possible scores). With reference to the English tests, this could be achieved by including more higher level questions in Lang1 (reading comprehension). However, Lang2 (Visual Literacy) would probably need different stimulus items altogether. The possibility of commissioning an artist to draw visual stimuli to specifications, would need to be explored. With reference to the Maths test, one means of extending the difficulty level of the test would be to include more items, thus making the test longer. However, a longer test might be too tiring for Grade 7 level learners, and this needs to be taken into consideration.
5. Significant discrepancies exist between the performance of learners at disadvantaged schools identified as "high-performing" and the performance of learners at SHC, which cannot be characterised as disadvantaged. SHC pupils at a Grade 6 level outperformed Grade 7 pupils at all the high-performing sample schools. This finding confirmed the expert opinion of the SHC Vice-Principal that the tests were indeed too easy.
6. Unlike assessment tools developed in other countries or elsewhere in South Africa, the St. Barnabas Tests are criterion-referenced, operationalising the principles of OBE and Curriculum 2005. In this respect, the St. Barnabas Tests differ significantly from tests in current use in schools which largely under-test significant aspects of C2005 (see Section 3.1.1). This test design means that the St. Barnabas Tests can be used to monitor the implementation of the new educational programmes in South Africa.
7. It can be concluded that for learner attainment in South Africa, it might not be possible to establish baseline criteria which are not context-specific. Disadvantaged ESL schools constitute a specific context, and learner attainment levels in other contexts can be expected to be very different. This confirms the point made in SAQA documentation:

"Consequently, interpreting and reporting learner achievement (outcomes) should be carefully contextualised. The results should always be interpreted against carefully selected input (resources) and process (learning environment) indicators, i.e. socio-economic background, physical amenities at the relevant institutions, disruption of learning opportunities, culture of learning, language of learning, etc." (Technical Committee of SAQA: Towards a Model for Assessment in Terms of an Outcome-based Approach. p. 42)

8. Although the St. Barnabas tests do not appear challenging enough, they have considerable diagnostic value regarding learner attainment in three different key areas of language

functioning and four different areas of Maths. Once modifications are undertaken to ensure that higher levels of performance can be better distinguished, the tests should be a valid (and norm-referenced) discriminator of learner attainment.

The St. Barnabas tests can be used in a variety of diagnostic situations:

- The tests could be used to evaluate the effectiveness of interventions (such as teacher development programmes) by focusing on learner attainment. This dimension of in-service teacher training programmes is often ignored. The St. Barnabas tests would serve the purpose of confirming that improved teaching had indeed resulted in better learner attainment.

- The individual components of the test could be used to identify strengths and weaknesses in instructional programmes at individual classrooms or schools. In this way, the tests contribute directly to progress towards identifying "best practices" in education.
- The tests could be used to identify strengths and weakness of individual learners. They could also be used for making recommendations regarding individualised instructional programmes for learners with special needs. This is in line with new proposals regarding "inclusion" and "mainstreaming" of learners with special needs. Many teachers feel inadequate as regards their ability to implement these proposals. Providing individualised diagnostic information on each pupil would assist teachers greatly.

7. FUTURE RESEARCH REQUIRED

The analysis conducted within this study is only the first step towards establishing the validity of the St. Barnabas instruments for use throughout South Africa. The following further tests and analyses will be required to confirm the utility of the tests.

- Strengthen the test components per the discussion above.
- Continue trial testing the instruments at other high-performing schools elsewhere in South Africa, especially among other language speakers so as to measure the effects of home language on performance. This would require standardisation of test administration procedures.
- Conduct item analyses (item difficulty analysis, item discrimination analysis, and distracter analysis) on individual questions in each test.
- Measure the influence of individual teachers and their training levels on learner attainment. If it is found that the achievement levels of pupils taught by particular teachers is significantly better than achievement levels of other pupils in a similar context, then research should be directed towards studying the classroom behaviour of the "high-performing" teachers. Such information would be invaluable in establishing "best practices" in education.
- Identify and analyse the educational practices which contribute to superior learner attainment in each of the skills assessed (e.g. perhaps time-on-task is the main contributor to superior performance in a skill?).
- Compare the test with other international and local tests.
- Measure the equivalency between the 3 Language elements (Language 1, 2, and 3) in each of the three test forms (Test A, B, and C).
- Conduct more detailed analysis of age effects on learner performance.
- Prepare an "item map" which links each test item to its relevant Specific Outcome under C2005.