Report to

The United Nations Children's Fund, Zimbabwe

from

The Australian Council *for* Educational Research

In collaboration with

Zimbabwe School Examinations Council

Evaluation of the Education Transition Fund Program

Zimbabwe Early Learning Assessment (ZELA)

2013 Monitoring Report

November 2013

Abstract

The Zimbabwe Education Transition Fund (ETF) program is a four-year project currently being undertaken by the United Nations Children's Fund (UNICEF). The ETF program provides essential material resources and support for the systems and structures designed to increase access to quality education for all Zimbabwean children. The ETF gives special attention to the most vulnerable children.

The Zimbabwe Early Learning Assessment (ZELA) project aims to improve Zimbabwe's system of pupil learning assessment. It does this by introducing an early-grades learning assessment to determine whether the EFT program (2010–2015) has worked. This includes producing positive outcomes for children, their care-givers, schools, and the education sector in general. The base-line is a measure of pupil performance in language and mathematics. Additional information was also collected by ZELA. This information was needed to more fully understand the performance of pupils. It included: pupil background characteristics, the availability of teaching resources, and the level of funding and facilities.

The tools developed in 2012 by ZELA provided a way of monitoring and evaluating the ETF program across the program's lifecycle. The key measure is the extent of improvement in pupil learning outcomes. The extent of this improvement is measured in scores on tests of language and mathematics in the early years of schooling.

UNICEF contracted the Australian Council for Educational Research (ACER) to undertake a base-line study in 2011, in collaboration with ZIMSEC. After the base-line study was completed in 2012, ACER was contracted to undertake two monitoring cycles (2013 and 2014) and an impact evaluation (2015), in collaboration with ZIMSEC.

Authorship

This report was prepared by Dr Rachel Outhred, Dr Gabrielle Matters, Timothy Chiwiye, Francis Chirume, Ebba Masiri, Samual Makore, Fredrick Mwale and Collet Mpofu.

Much of the analyses undertaken to inform this report were conducted by a team of ZIMSEC Specialists as part of an SPSS workshop, conducted by ACER as part of ZELA Capacity Building activities. In August 2013 Louise Wenn and Rachel Outhred facilitated a workshop with ZIMSEC colleagues; Timothy Chiwiye, Francis Chirume, Ebba Masiri, Samual Makore and Fredrick Mwale. Topic included sampling and standard errors, correlations and regressions and IRT and plausible values. In October 2013 Jorge Fallas and Rachel Outhred facilitated a workshop where ZIMSEC personnel applied their data analysis skills to the ZELA 2013 data. ZIMSEC staff ran analyses included in this report including calculating the mean performance of population subgroups and determining if differences are statistically significant, calculating confidence intervals for the mean performance of population subgroups within the sample, creating graphs and tables including error bar charts, bar charts, line graphs, pie charts, population pyramids, scatter plots and histograms, creating cut points in line with Standards, calculating the percentage of pupils within each band and analysing pupil performance against pupil and school characteristics.

This report draws on the contributions of the ZELA Project Team, which include other key staff at ACER and ZIMSEC– including but not limited to Timothy Chiwiye, Francis Chirume, Ebba Masiri, Samual Makore, Fredrick Mwale, Collet Mpofu, Eveline Gebhardt, Jorge Fallas, Louise Wenn, Dr Mark Wall, Martin Murphy and Chris Freeman. Dr Gerald Elsworth of Deakin University, Melbourne undertook the multilevel analyses.

Abbreviations and acronyms

ACER	Australian Council for Educational Research
BEGE	Basic Education and Gender Equality
BESO	Basic Education System Overhaul
ECD	Early Childhood Development
EFA	Education for All
ETF	Education Transition Fund
IRT	Item Response Theory
MC	Multiple Choice
MDG	Millennium Development Goal
MLA	Monitoring Learning Achievement
NGO	Non-Government Organization
OVC	Orphans and Vulnerable Children
PD	Professional Development
PISA	Programme for International Student Assessment
TMO	Test Monitoring Officer
SACMEQ	Southern Africa Consortium for Monitoring Educational Quality
UNICEF	United Nations Children's Fund
ZELA	Zimbabwe Early Learning Assessment
ZimSEC	Zimbabwe Schools Examination Council
ZimVAC	Zimbabwe Vulnerability Assessment Committee

Contents

Executive Summary

The Zimbabwe Early Learning Assessment (ZELA) Education Transition Fund (ETF) program is a four-year project commissioned by the United Nations Children's Fund (UNICEF) to support and enhance national capacity to review, reform, and re-orientate the current system of pupil assessment in Zimbabwe and to establish a baseline for determining whether the ETF program (2010–2015) has the desired effects on children, their care-givers, schools, and the education sector in general, as well as to identify the extent to which changes identified are attributable to the ETF program interventions.

In 2010 UNICEF and its partners supplied all Zimbabwean schools with resources as part of the ETF, aimed at attaining quality education and access for all. The intervention includes the provision of textbooks and the establishment of supporting infrastructure relating to the use of these books in schools.

This report relates to Phase Two of the ETF Learning Assessment Study and Evaluation. This phase is concerned with building on the baseline data and information collected n 2012 for the evaluation of the ETF program.

The major research questions for the study are:

1. How do the Zimbabwe pupils perform in the language and mathematics tests? Closely related to this is the question: Is there a noticeable pattern of change over time?
2. What are the relationships of the following groups of variables with performance on tests of language and mathematics at the beginning of Grade 3 in Zimbabwe?
    - Pupil background characteristics
    - Teacher and teaching resources
    - School funding and facilities
3. To what extent can improvement in test performance be attributed to the Education Transition Fund?

Sample and Data Collection

The target population was pupils in Grade 3 in Zimbabwe. A representative sample was drawn which yielded approximately 16 000 pupils in about 500 schools across the ten provinces of Zimbabwe.[1]

Four ZELA tests were administered, namely English, Ndebele, Shona and Mathematics.

Responses to the Pupil Questionnaire, which contained 18 questions, were collected from 13 271 Grade 3 pupils. Responses to the School Head Questionnaire, which contained 40 questions, were collected from 411 School Heads (or their delegates).

[1] For further information on the sampling process please see the ZELA 2013 Technical Report.

ZELA Tests

Each test is aligned to the Grade 2 curriculum in Zimbabwe and administered at the beginning of Grade 3. A single scale aligning the abilities of pupils with the difficulties of the items was constructed for each ZELA test (namely English, Ndebele, Shona and Mathematics) after thorough test-by-test analyses. For each scale (English, Ndebele, Shona and Mathematics), the distribution of pupil abilities in ZELA 2012 was transformed to a scale with a mean of 300 and a standard deviation of 25.[2] Link items from the 2012 test were used in the 2013 tests. All items were calibrated separately for each of the four cognitive domain in order to ensure the 2012 and 2013 test results were comparable.[3]

Findings

1. How do the Zimbabwe pupils perform in the language and mathematics tests? Closely related to this is the question: Is there a noticeable pattern of change over time?

Grade level benchmarks

The 2012 base-line study found that for **English**, 49.2 per cent of Grade 3 pupils were achieving at or above the grade level benchmark, and that for **Mathematics**, 45.8 per cent of pupils were achieving at or above the grade appropriate level. In 2013 there was a 4.5 per cent increase in the number of Grade 3 pupils achieving at or above the grade appropriate level in Zimbabwe. In 2013 there was a 17.1 per cent increase in the number of pupils achieving at or above the grade appropriate level.

For the African languages that were used in the tests, it was found that;

- In 2013 78.2 per cent of the pupils who completed Shona tests were achieving at or above the grade-appropriate level in **Shona** after completing Grade 2 in Zimbabwe
- In 2013 78.5 per cent of the pupils who completed Ndebele tests were achieving at or above the grade-appropriate level in **Ndebele** after completing Grade 2 in Zimbabwe.

Changes in average performance

The average student in performance improved across all tests between 2012 and 2013 as in 2012. There were statistically significant increases in pupil performance across language tests (English, Shona and Ndebele[4]) and the Mathematics test. Increases in performance were similar for boys and girls.

Mashonaland Central was the only province where a statistically significant increase in pupil performance in both English and Mathematics was observed. Significant increases in Mathematics

[2] For further information on test reliability and the psychometric analysis of the tests, please see the ZELA 2013 Technical Report.

[3] For further information on the calibration process, please see the ZELA 2013 Technical Report.

[4] There were increases in pupil performance in both Shona and Ndebele for those pupils undertaking the African language tests. However, the sampling was not undertaken to achieve a nationally representative sample of Shona and Ndebele speakers, so the 2012 and 2013 results are not directly comparable

were observed in Manicaland, Mashonaland Central, Mashonaland East, Masvingo, Matabeleland South and Midlands.

There was no evidence to suggest that the performance of pupils from Satellite schools had increased.

2. What are the relationships of various variables with performance on tests of language and mathematics at the beginning of Grade 3 in Zimbabwe?

Of the range of pupil background characteristics that were investigated, socio-economic status was found to be one of the strongest predictors of pupil achievement levels in Zimbabwe. On average, pupils in the highest SES quartile were the only group achieving above the national mean.). Indeed, there were statistically significant differences in student performance between all SES quartiles. This indicates that the learning outcomes of pupils are being impacted on by SES.

The textbook pupil ratio in urban areas had a weak to moderate negative association with performance in urban areas only (as the number of pupils sharing textbooks increases, the performance of pupils decreases). There were no or weak and insignificant associations with textbook ratio and pupil performance in rural areas. The percentage of qualified teachers had a weak, yet significant negative association with pupil performance in urban areas and a weak to moderate, yet significant association in rural areas. This indicates that as the percentage of qualified teacher's increases in urban schools, the performance of pupil's decreases. This could be due to a range of issues, including poor teaching training, hiring highly educated, yet unqualified teachers in urban areas or could simply be due to the high percentage of teachers who are qualified in urban schools. In rural schools the higher the percentage of qualified teachers, the higher pupil's performed. The number of school days lost during the first term of 2013 did not appear to have an association with pupil performance and teacher absenteeism did not have a strong association with performance. This was probably due to the low levels of school closures and teacher absenteeism.

The predicted performance[5] of pupils rose steadily between the per pupil budget ranges of US$50 to US$175. The analysis found that below this range different budget levels are unlikely to impact on pupil performance. Above it, while the predicted pupil performance was higher, increases in pupil budget were less likely to further increase performance. In other words, there appears to be two thresholds below and above which the level of budget allocations are less likely to result in impacts on pupil performance. Weak to moderate associations between pupil performance and school access to piped, tank or spring water and electricity were observed.

3. To what extent can improvement in test performance be attributed to the Education Transition Fund?

The indicator currently used to monitor pupil progress is the percentage of pupils achieving at a grade appropriate level. This logically signals the ETF Program intervention target population to be those pupils below the grade-appropriate level and schools and provinces with the greatest percentage of pupil's achieving below this level.

Analysis of the UNICEF distributed textbooks indicates that the textbook distribution has appropriately targeted these pupils, communities and provinces. Provinces with the lowest average

[5] Please see the ZELA 2013 Monitoring Report for full details of the regression analysis undertaken.

pupil performance have a greater number of UNICEF distributed textbooks per pupil (on average). At the pupil level, the data also indicates that pupils achieving below the grade appropriate level have, on average, more UNICEF distributed textbooks than pupils achieving at or above the grade appropriate level.

Two hypotheses suggested by the analyses conducted in 2012 included:

1. An outcome of the intervention will be an increase in average achievement over the time of the intervention; and,
2. An outcome of the intervention will be a reduction in this proportion of school-level variance in achievement.

There have been increases in pupil performance in English and Mathematics across the majority of provinces in Zimbabwe. The patterns of achievement by province and gender are very similar for English and Mathematics (however the gap between urban and rural performance for English is greater than for Mathematics). The majority of provinces had a positive increase in performance in Mathematics and English and no provinces had a statistically significant decline in performance. Increases in Mathematics performance were greater than English performance, however the improvements made to the tests in 2013 may also have had an impact on the equating process (this can be determined in 2014).

The school-level variance in performance was found to be relatively high, indicating that which school a pupil attends has a large effect on their performance. In line with the aims of the ETF Program, we would expect to see a reduction in the proportion of school level variance over the ETF Program cycle. The proportion of school-level variance in achievement will continue to be monitored and reported on over the 2014 ZELA monitoring cycle and will be used to determine the extent to which changes in pupil performance can be attributed to the ETF Program. Making this attribution is one of the main aims of the wider ETF evaluation study.

Chapter 1 Education Transition Fund and the Right to Education

Basic Education and Gender Equality (BEGE) is a key focus within UNICEF's Medium-Term Strategic Plan, which outlines UNICEF's contributions to the Millennium Development Goals (MDGs). Since 1946, UNICEF has worked to ensure the rights of all children and, guided by the Convention on the Rights of the Child, the Millennium Declaration and Education for All, this has included the right to education.

The UNICEF Basic Education and Gender Equality (BEGE) Focus Area supports national capacity to improve children's developmental readiness, reduce disparities in education, improve educational quality, retention, completion and achievement, and restore education in emergencies and post-conflict situations.

UNICEF's 'leadership agenda for equity' sharpens the strategic focus of its programs and policies to place an even greater emphasis on the most disadvantaged children, through supporting national capacity for scaling-up interventions, providing technical and programmatic guidance and engaging in strategic partnerships and policy dialogue.

In partnership with national governments, civil society actors, and other development partners, the Education Section maintains a commitment to work towards the provision of quality basic education for all children globally, and to enhance support for education for marginalized groups.

1.1 Scope of the study

The major research questions for the study are:

1. How do the Zimbabwe pupils perform in the language and mathematics tests? Closely related to this is the question: Is there a noticeable pattern of change over time?
2. What are the relationships of the following groups of variables with performance on tests of language and mathematics at the beginning of Grade 3 in Zimbabwe?
    - Pupil background characteristics
    - Teacher and teaching resources
    - School funding and facilities
3. To what extent can improvement in test performance be attributed to the Education Transition Fund? Addressing this question will need to take account of different levels of other resources and facilities in the schools.

A subset of these research questions relates to the base-line study, the focus of this report. The questions that the current study aims to answer are:

- How do early-grade Zimbabwe pupils perform in tests of language and mathematics?
- Is it possible to identify pupil-level and school-level variables that influence test performance?

In order to answer these questions, the following was done:

- There was a review of the literature on international experience in national assessment with particular emphasis on Africa. This was done in order to ensure that the project worked with the latest and best information for implementing the project. The review of literature in documented in the ZELA 2012 Base-line Study report.
- A representative sample of Grade 3 pupils from schools across the ten provinces of Zimbabwe was drawn in 2012 and 2013 using the same sampling strategy. This sampling strategy allows the study to generalise to the population of all Grade 3 pupils in Zimbabwe.
- Tests of Mathematics, English, Ndebele, and Shona were developed and administered in 2012 and 2013.
- Security procedures surrounding test development, printing, administration and marking were developed and implemented.
- Questionnaires was developed and administered in 2012 and 2013. This was designed to collect information about pupil backgrounds and the school and the teachers at the school.
- A manual for school administrators was developed to ensure the tests were administered consistently and appropriately.
- School administrators were trained to ensure that they fully understood test protocols and the reasons for them.
- Test Monitoring Officers (TMOs) were appointed and trained. These officers were important for ensuring that the quality of the data was protected at key stages during their collection and processing.
- Training manual for school administrators to TMOs was produced. These provided guidelines for quality assurance practices as well as being the basis for their feedback to ACER on the conduct of the tests.
- Visits to schools by ACER staff and TMOs were conducted as part of the quality assurance for the study.
- Guides to school heads, teachers and parents were developed and dispatched.
- Procedures for data capture – that is, moving the information from the completed test forms and questionnaires to an electronic format.
- Data entry (in Zimbabwe) and data cleaning (in Australia)

At this point the data analysis and report writing commenced.

1.2 The implementation of the study over time

The progression of cohorts of pupils being surveyed is summarised in Table 1. The number in parenthesis refers to the order of the test cycles over the duration of the evaluation. This report addresses the 2013 cycle. 2013 is the second cycle of ZELA, and is the first monitoring cycle after the initial base-line study.

Table 1 Location of the 2013 cycle within the wider evaluation

2010	2012	2013	2014	2015
Grade 2 (1)	Grade 2 (2)	Grade 2 (3)		
EGLALN	Grade 3 1st cycle	Grade 3 2nd cycle	Grade 3 3rd cycle	Grade 3 4th cycle
NA	Base-line	Monitoring	Monitoring	Evaluation

1.3 Limitations of the study

In 2013 the tests were improved from 2012, particularly the lay-out and design. This improvement brought risks to the equating process. The tests were equated in order to compare the performance of pupils between 2012 and 2013. However, changes in the average mean abilities reported in this document may be in part attributable to changes in the position of link items and changes in the lay-out of the tests. The extent to which this is the case will be able to the determined in the next ZELA round.

As with the base-line study, the mathematics test was set in English and attempts were made to minimise the reading load on pupils. Nevertheless, it can be expected that mathematics is not the only construct being assessed, and that an irrelevant construct (such as reading comprehension) is also present in the mathematics tests. This phenomenon will occur in tests of mathematics that are set for pupils for whom English is their first (in some cases, only) language. Its effect will probably be stronger in a test of mathematics set for pupils for whom English is their second language.

The administration of ZELA 2012 and ZELA 2013 was at the beginning of Grade 3 rather than at the end of Grade 2, and the test was curriculum-based. While the test developers aimed to use curriculum-appropriate content, and to pitch the test at the appropriate level for the pupils with respect to item difficulty, setting the test at the start of the school year has limitations. Pupils tend to regress in their achievement levels over holiday periods. They are not well established in the school routine and learning and so testing may not be undertaken in stable settings, which may impact on performance. Quite where to pitch the level of the tests, and what curriculum to reference are thus made more difficult. The 2013 test was administered three weeks later in the school term than the 2012 test. This was due to unavoidable delays. This resulted in a slightly higher rate of pupils attending only one of the two days of test administration due to absenteeism in the last week of the school term. Given the large sample this did not present significant issues.

1.4 Summary of Chapters

This chapter (Chapter one) introduces the Evaluation of the Education Transition Fund Program, including a description of the scope of the study, how the study is being implemented over time and the limitations of the study. Chapter two addresses the first research question: *How do the*

Zimbabwe pupils perform in the language and mathematics tests and is there a noticeable change in performance over time? The chapter provides a picture of achievement in Zimbabwe at the national level and by sub-groups including gender, location, province, school type and age and describes the changes observed in pupil performance since baseline. Chapter three reports on the predictors of performance on the tests of language and mathematics including pupil background characteristics, teacher and teaching resources and school resources (funding and facilities). Chapter four address the extent to which observed changes can be attributed to the Education Transition Fund Program. Chapter five concludes with future directions for the monitoring of pupil achievement in Zimbabwe and proposes a number of policy implications. The theoretical framework for the study, indicative standards by proficiency level, project documentation and methodology are provided as Annexes to the report.

Chapter 2 – How do the Zimbabwe pupils perform in the language and mathematics tests?

This chapter addresses the research question: *How do early-grade Zimbabwean pupils perform in tests of language and mathematics?* And the related question: *Is there a noticeable pattern of change over time?*

It first provides a description of differences between various subgroups. It then examines the proportion of pupils below, at and above grade level benchmarks and reports on changes in performance between 2012 and 2013.

2.1 Sub-group differences in English and Mathematics

2.1.1 Gender

Girls outperformed boys on average in English and Mathematics, but only slightly. See Table 2.

Table 2 Summary statistics showing scores in English for boys and girls, 2013

	Gender	Mean	N	Std. Deviation
English 2013	Boys	302.0	6208	23.5
	Girls	306.6	6741	23.9
Mathematics 2013	Boys	308.5	6208	22.1
	Girls	311.8	6251	23.8

2.1.2 Location

Pupils in urban schools significantly outperformed pupils in rural schools in both English and Mathematics.

Table 3 Summary statistics showing scores in English and Mathematics for rural and urban pupil, 2013

	Location	Mean	N_cases	Std. Deviation
English 2013	Rural	297.2	12392	19.1
	Urban	326.2	3586	25.0
Mathematics 2013	Rural	305.7	12392	20.5
	Urban	324.2	3586	22.0

2.1.3 Gender differences by province

The Shona and Ndebele tests were not administered in all provinces.

Figures 1 and 2 display the mean scores on English and Mathematics, respectively. The value of the mean is denoted by the location of a circle in the graphs. For example, the mean score for English by boys in Bulawayo is around 317 and this can be seen in circle located furthest to the left in this figure. The 95 per cent confidence interval is shown by the vertical line running through the mean. This interval provides an indication of where the true mean is most likely to be located. (Because the study uses a sample, the means shown are the best estimates available based upon the sample, which may vary from the true population mean).

The patterns of achievement by province and gender are very similar for English and Mathematics, however the gap between urban and rural performance for English is greater than for Mathematics. This is likely to reflect the lower proportion of pupils exposed to English in rural locations.

Figure 1 English scale score (ESS) means, by province and gender

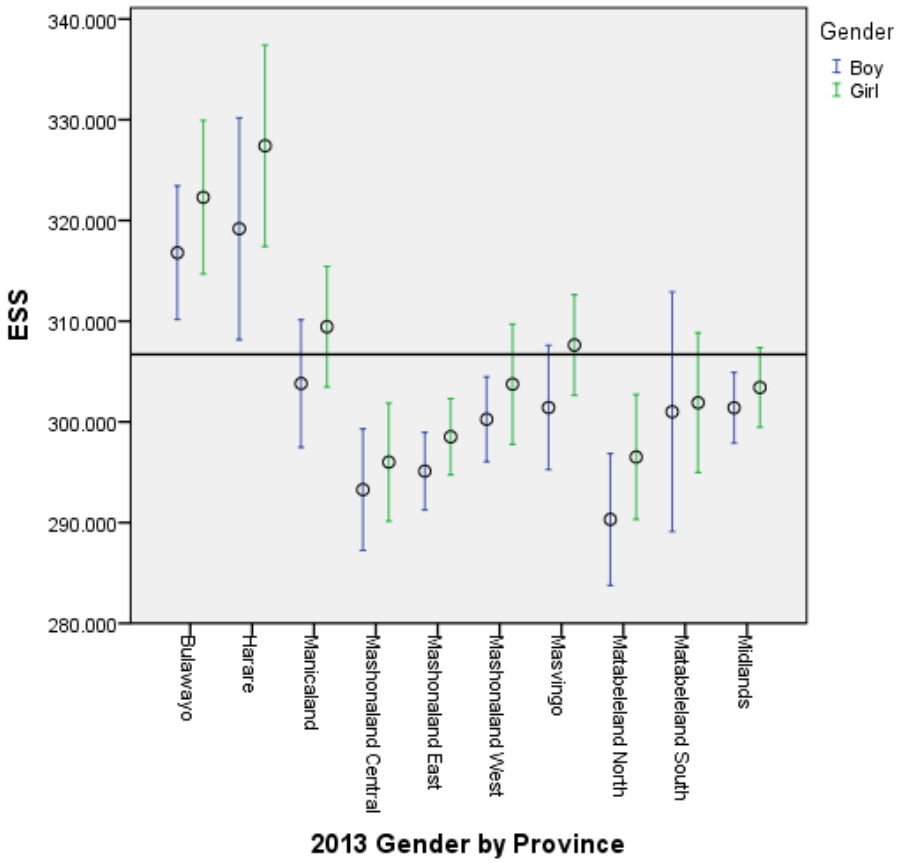
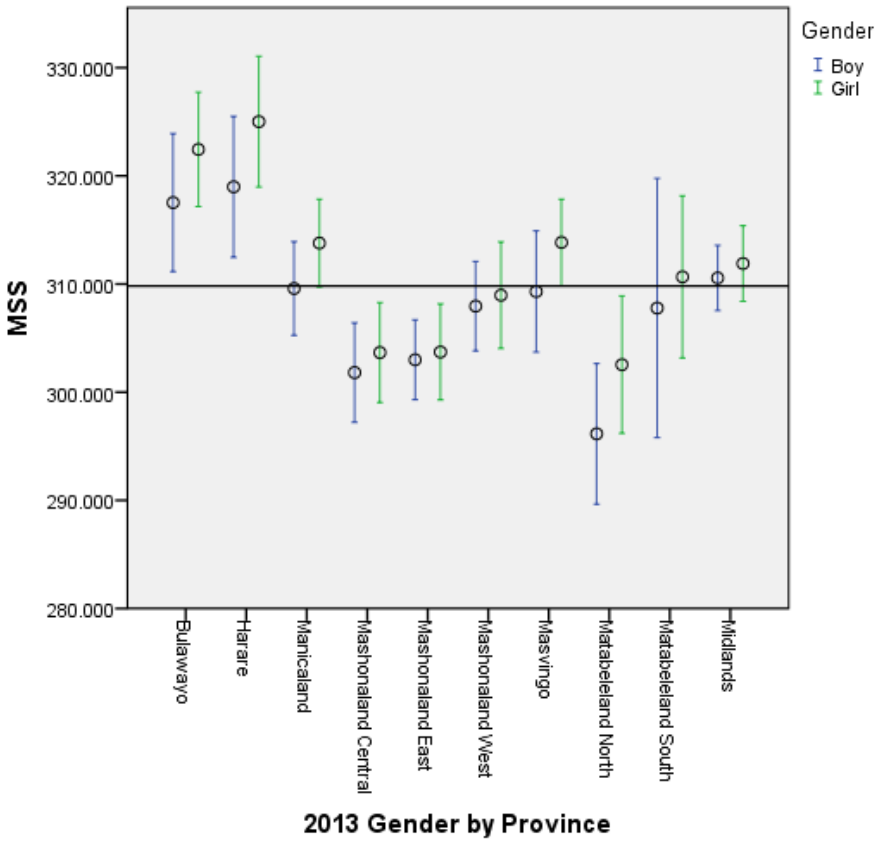


Figure 2 Mathematics scale score (MSS) means, by province and gender



2.1.4 School Type

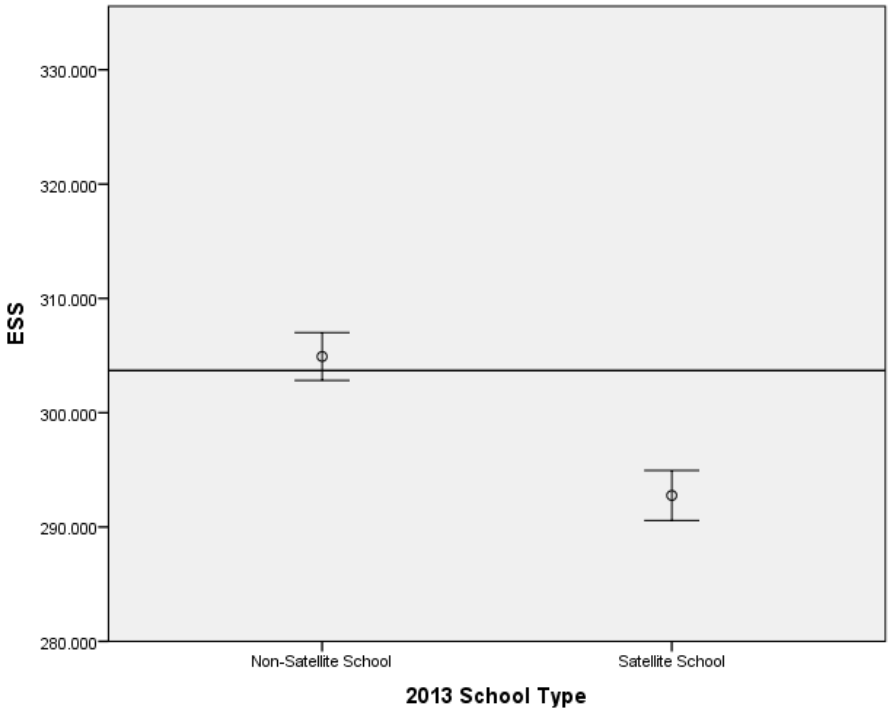
Pupils from non-satellite schools outperformed pupils from satellite schools on average in English and Mathematics by a large margin.

Test	Location	Mean	N_cases	Std. Deviation
English 2013	Non-Satellite	304.92	14372	24.0
	Satellite	292.8	1606	19.4
Mathematics 2013	Non-Satellite	310.6	14372	22.5
	Satellite	302.9	1606	19.2

Table 4 Summary statistics showing scores in English and Mathematics for pupils within satellite and non-satellite schools

As can be seen in the figure below, the evidence indicates the average score of pupils in Satellite schools is well below that of pupils in non-Satellite Schools.

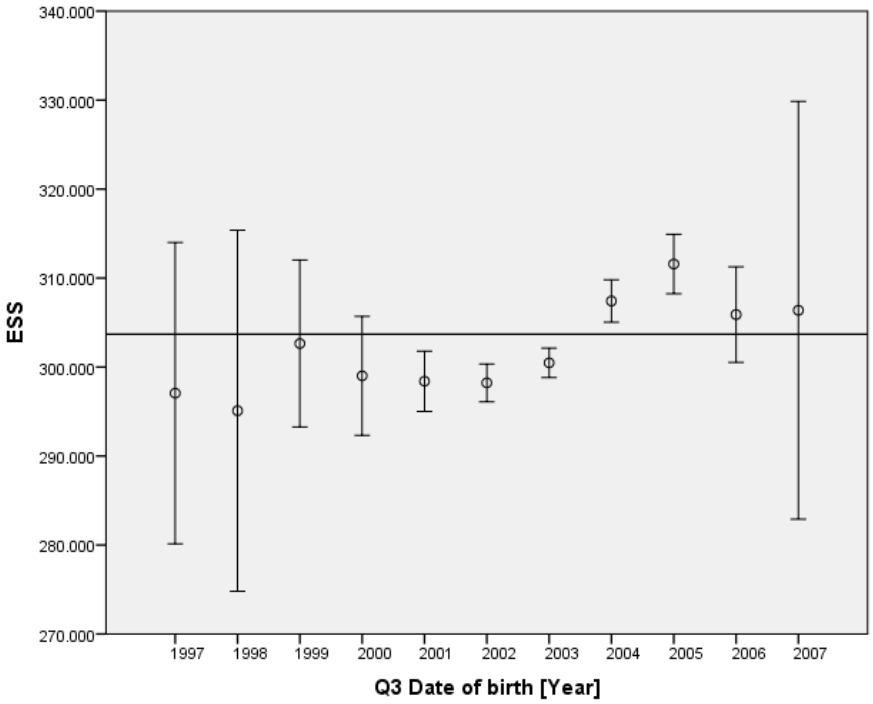
Figure 3 English scale score (ESS) means by school type



2.1.5 Age

Pupils born in 2004 and 2005 (approximately seven or eight years of age) performed better on the English and Mathematics tests, than pupils born in the four years previous.

Figure 4 English scale score (ESS) means by Age



2.2 Is there a noticeable pattern of change in pupil performance over time?

At the national level, statistically significant increases in pupil performance were observed across language tests (English, Shona and Ndebele) and the Mathematics test between 2012 and 2013.

2.2.1 English

The mean performance of pupils on the English test increased since 2012. All of the provinces, with the exception of Mashonaland Central had a positive increase. However, the only province where the increase was statistically significant was Manicaland. There was no evidence to suggest that the performance of pupils from Satellite schools had increased significantly.

2.2.2 Mathematics

The mean performance of pupils increased since 2012. All of provinces had a positive increase. Statistically significant increases included Manicaland Mashonaland Central, Mashonaland East, Masvingo, Matabeleland South and Midlands.

As can be seen in **Error! Reference source not found.** and Figure 6, the increases in performance were even for boys and for girls, with girls outperforming boys, as in 2012.

Figure 5 English scale score (ESS) mean increases by Gender, 2012 and 2013

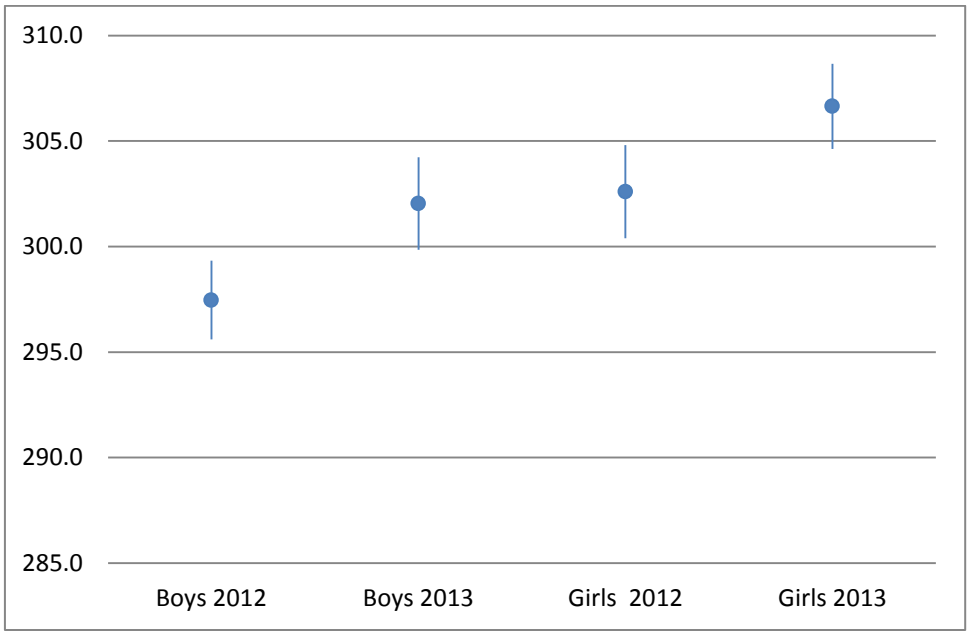
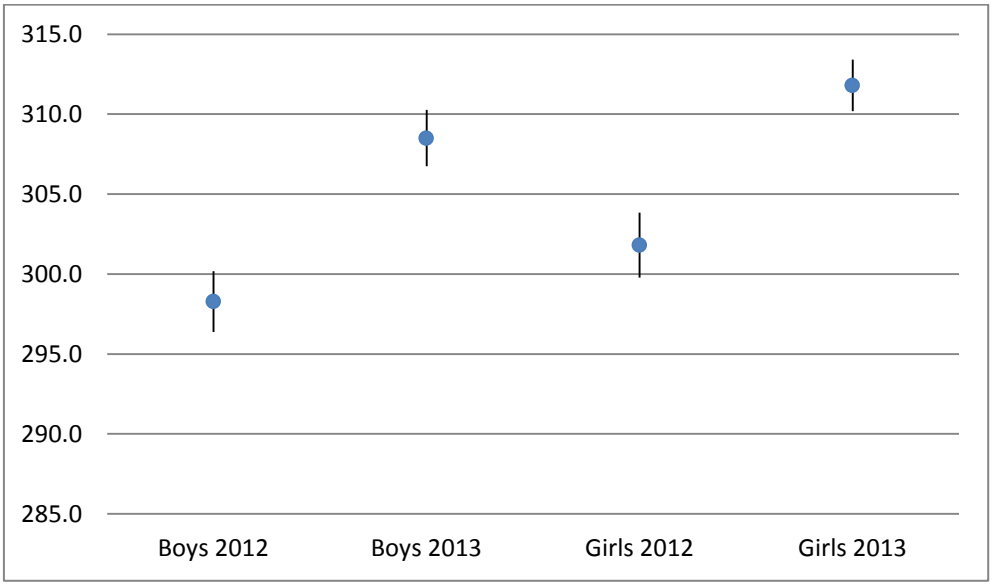


Figure 6 Mathematics scale score (MSS) mean increases by Gender, 2012 and 2013



As can be seen in Figure 7 and Figure 8, increases in pupil performance in English were not significant between 2012 and 2013 for rural or urban subgroups, but were significant between the subgroups for Mathematics.

Figure 7 Mathematics scale score (MSS) mean increases by Gender, 2012 and 2013

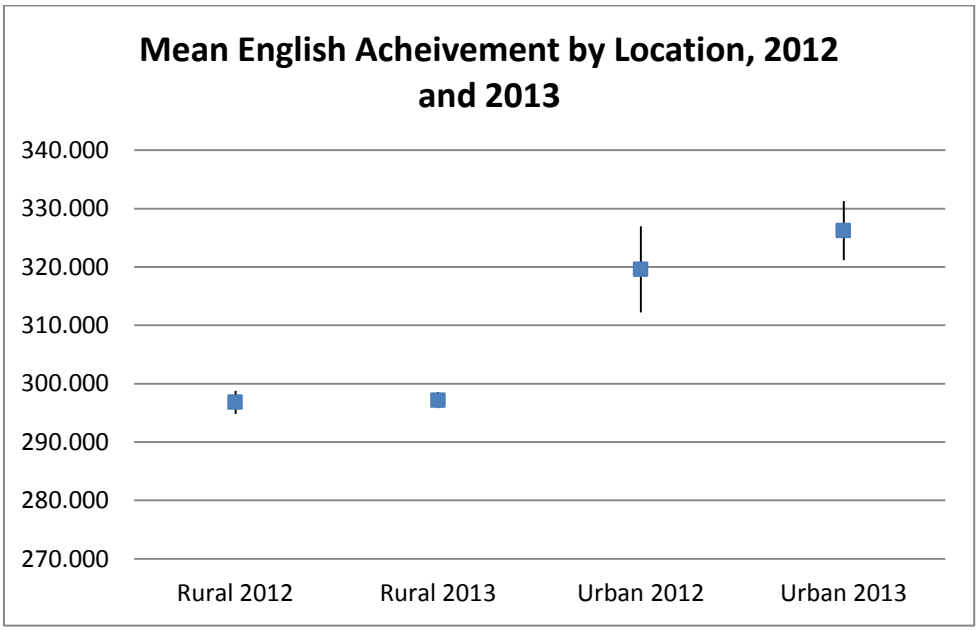
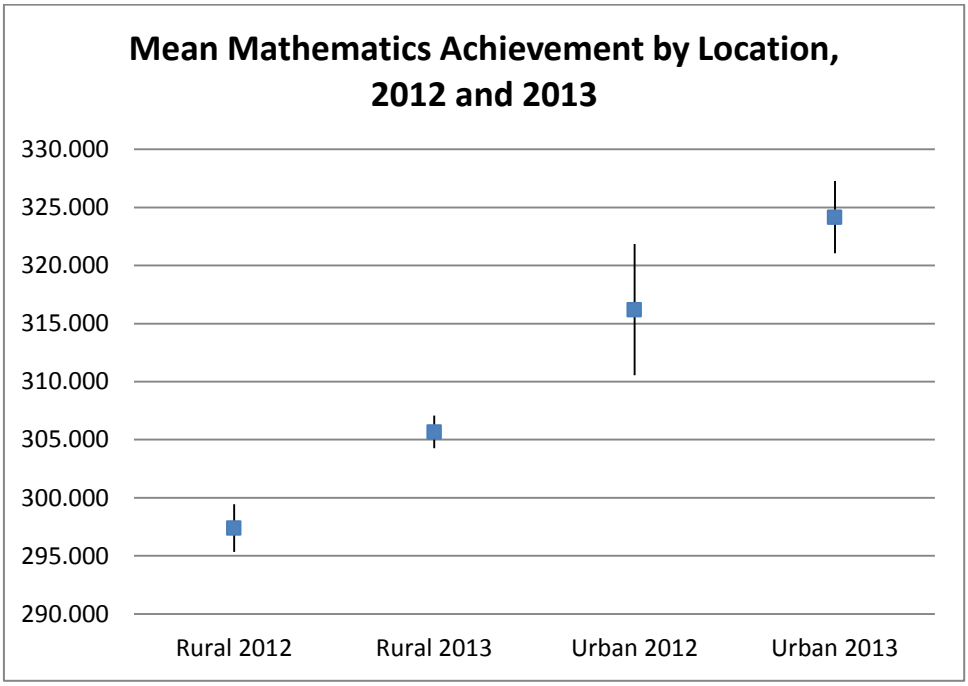


Figure 8 Mean Mathematics Achievement by Location, 2012 and 2013



2.2.3 School Type

Increases in pupil scores for Mathematics were significant for Satellite schools. However, pupil performance in English was not significant for pupils from Satellite schools.

Test	Location	Mean	N_cases	Std. Deviation
English 2012	Non-Satellite	300.8	15975	25.4
	Satellite	291.4	1502	18.7
English 2013	Non-Satellite	**304.92****	14372	24.0
	Satellite	292.8	1606	19.4
Mathematics 2012	Non-Satellite	300.6	15975	25.2
	Satellite	293.4	15975	21.2
Mathematics 2013	Non-Satellite	**310.6****	14372	22.5
	Satellite	**302.9****	1606	19.2

***** Statistically significant increase from 2012***

Table 5 Summary statistics showing scores in English and Mathematics for pupils within satellite and non-satellite schools

2.2.4 The African Languages

There were increases in pupil performance in both Shona and Ndebele for those pupils undertaking the African language tests. However, as the sampling was not undertaken to achieve a nationally representative sample of Shona and Ndebele speakers, the results are not directly comparable.

Increases in Shona performance were significant in two of the seven provinces where Shona tests were administered (Harare and Midlands). Increases in Ndebele performance were significant in two of the three provinces where Ndebele tests were administered (Bulawayo and Matabeleland South). Increases in Shona performance for pupils in Satellite and non-Satellite schools were observed, however Ndebele performance in Satellite schools did not significantly increase.

2.3 Proportion of pupils at, above and below the grade level benchmarks

The indicator used to measure changes in pupil achievement in English and Mathematics in Zimbabwe is the percentage of pupils achieving at or above the grade-appropriate level after completing Grade 2 in Zimbabwe.

2.3.1 Achieving at a level

'Achieving at a level' means a pupil would get at least half of the items correct in a test set specifically for pupils at their grade level. This test should be composed of questions ranging in difficulty, but always appropriate to the year level of the pupils. So some questions would be easier than others, but they would not fall outside this range. The cut-point separating 'below' the grade-appropriate level and the 'at' grade-appropriate level represents the minimum acceptable standard in the subject.

2.3.2 Proportion of pupils achieving below, at and above the grade-appropriate level in English

The 2012 base-line study found that the percentage of pupils performing at or above the grade-appropriate level after completing Grade 2 in Zimbabwe is 49.2 per cent for English. The 2013 monitoring cycle found that the percentage of pupils performing at or above the grade-appropriate level after completing Grade 2 in Zimbabwe in 53.6. This represents a 4.5 per cent increase in the per cent of pupils achieving at or above the grade appropriate level in one year. See Table 6.

Table 6 The proportion of pupils below, at and above the grade level benchmark for English

	2012 (Per cent)	2013 (Per cent)	Per cent increase 2012 - 2013
Below grade level	50.8	46.4	-4.5[6]
At grade level	39.4	41.1	1.7
Above grade level	9.8	12.5	2.7
At and above grade level	49.2	53.6	4.5

[6] Note: The per cent increase is -4.5% rather than 4.4% due to the rounding of percentages to one decimal point in columns B and C.

2.3.3 Proportion of pupils below, at and above the grade level benchmarks for mathematics

The 2012 base-line study found that the percentage of pupils performing *at or above* the grade-appropriate level after completing Grade 2 in Zimbabwe is 45.8 per cent for Mathematics. Thus, 54.2 per cent were below the appropriate level. See Table 7. The 2013 monitoring cycle found that the percentage of pupils performing *at or above* the grade-appropriate level after completing Grade 2 in Zimbabwe is 62.9 per cent for mathematics. Thus 37.1 per cent were below the appropriate level. This represents a 17.1 per cent increase in performance on mathematics tests. In 2012 the distribution of pupil's performance peaked (just) below the 'at grade level' cut-point, indicating a large number of pupils were on the cusp of achieving the grade-appropriate level. This is one likely explanation for the large percentage increase in mathematics performance. However, the improvements made to the tests in 2013 may also have had an impact on the equating process. The 2014 ZELA cycle will allow judgements to be made regarding the extent to which the increase in performance can be attributed to the improved test format.

Table 7 The proportion of pupils below, at and above the grade level benchmark for mathematics

	2012 (Per cent)	2013 (Per cent)	Per cent increase 2012 - 2013
Below grade level	54.2	37.1	-17.1
At grade level	36.2	46.6	10.5
Above grade level	9.6	16.3	6.6
At or above grade level	45.8	62.9	17.1

2.3.4 Proportion of pupils below and at or above the grade level benchmarks in the African languages

In acknowledgement that some pupils have not yet had the opportunity to learn from the English syllabus, pupils were given the option to complete a test in Shona or Ndebele, in addition to English and Mathematics tests. As stated above, in both 2012 and 2013 pupils were not randomly sampled from the population group for this activity; therefore their performance on the test is not representative of the population group. Further, the Shona and Ndebele tests were constructed from the English test so the Shona and Ndebele tests items draw on language rules applied from English to Shona and from English to Ndebele. The full range of principles and processes by which sentences are structured and meaning is made of text in Shona and Ndebele were not, therefore, tested.

For a number of pupils who completed the Shona or Ndebele test, the test was in the native language of that pupil, however for others it was not. Therefore, the Shona and Ndebele tests were measuring neither literacy nor language acquisition across the sample. Furthermore, the three language tests are not comparable. Therefore pupil achievement in Shona cannot be compared to pupil achievement in Ndebele or to pupil achievement in English.

Collectively, these constraints translate into test results that cannot be considered representative of the national situation. Therefore, results should be considered indicative of the extent to which a sample of pupils in Zimbabwe has mastered a sample of the principles that are applicable to the languages of Shona and Ndebele.

Based on the application of cut-points developed by subject experts from ZIMSEC:

- In 2012 63.7 per cent of the pupils who completed Shona tests were achieving at or above the grade-appropriate level in **Shona** after completing Grade 2 in Zimbabwe
- In 2013 78.2 per cent of the pupils who completed Shona tests were achieving at or above the grade-appropriate level in **Shona** after completing Grade 2 in Zimbabwe
- In 2012 66.6 per cent of the pupils who completed Ndebele tests were achieving at or above the grade-appropriate level in **Ndebele** after completing Grade 2 in Zimbabwe
- In 2013 78.5 per cent of the pupils who completed Ndebele tests were achieving at or above the grade-appropriate level in **Ndebele** after completing Grade 2 in Zimbabwe.

In 2012 the majority of pupils in Zimbabwe were achieving below the grade appropriate level. However, in 2013 increases in performance have resulted in a large minority of pupils achieving below the grade appropriate level. In 2014 judgements will need to be made, regarding the extent to which increases in performance were the result of the improvement of the test lay-out and design. Therefore, the 2013 percentages of pupils achieving at or above the grade appropriate level may need to be retrospectively adapted in 2014. With that said, given the distribution of pupil performance peaked just below the *'at grade appropriate'* cut-point in 2012, and the rapid increases in advancements that can be observed when pupils start to read; it is possible the improvements are largely real.

2.4 Conclusions

The observed differences in pupil performance on the tests of language and mathematics between sub-groups in Zimbabwe are similar to those from other studies in many countries regarding gender and location (urban versus regional) and observed differences in pupil performance in Zimbabwe in 2012.

In Southern Africa differences in gender performance are varied, with boys performing better in some school systems, and girls performing better in others.[7] As stated in the 2012 baseline report, it is important to consider, not just differences in performance, but also the direction and degree of differences in the system over time. At this very early stage of monitoring the performance of pupils in the early years in Zimbabwe, trends appear to be following that of other Southern African nations.

While at the national level increases in pupil performance have been observed, the low and/or insignificant increases in English performance in Satellite schools and rural schools translates into wider achievement gaps within Zimbabwe. Government, donors and partners are implementing programs and policies to address the achievement of pupils in Satellite schools and in rural

[7] Trends in the Magnitude and Direction of Gender Differences in Learning Outcomes, SACMEQ Working Paper September 2011. P11.

schools, the results of which will become clearer as the monitoring of the performance of Zimbabwean pupils continues as part of the ETF program evaluation.

Chapter 3- Predictors of performance on tests of language and mathematics at the beginning of Grade 3 in Zimbabwe

This chapter addresses the second research question of the study: *What are the relationships of the following groups of variables with performance on tests of language and mathematics at the beginning of Grade 3 in Zimbabwe?* It considers:

- Pupil background characteristics
- Teacher and teaching resources
- School funding and facilities

3.1 Pupil background characteristics

This section examines the relationship between test scores and various pupil background characteristics. These include:

- Socio-economic status
- Language of instruction
- Language spoken at home
- Number of books in the home
- Time spent working for the family
- Whether the pupil attended an early Childhood Development Class
- Levels of parental education
- Pupil absence

3.1.1 Socioeconomic Status

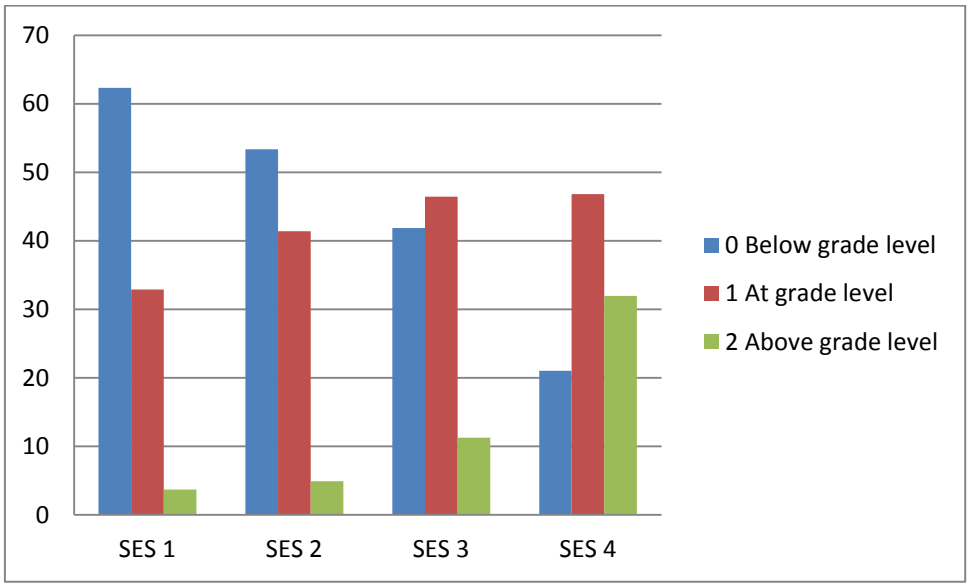
An index of socioeconomic was developed for 2012 and 2013. Five components were used including books at home, meals per day, highest parental education, home possessions and home educational resources.[8] The socioeconomic index was then used to determine the extent to which the variance in pupil performance can be attributed to socioeconomic status. Analyses were undertaken using both the fine grain SES Index variable and in addition Grade 3 pupils were categorised into SES quartiles. The analysis found that the mean score of pupils within the highest SES quartile was 26.5 scale scores above the mean score of pupils in the lowest SES quartile for English and 20.6 scale scores above the mean score in of pupils in the lowest SES quartile for Mathematics.

The difference in performance by SES was significant between all SES quartiles, indicating that there is a real difference between the learning outcomes of pupils based on SES.

[8] See the ACER Technical report for details on how the SES variable was developed

Figure 9 represents the percentage of pupils achieving below, at and above grade level per SES quartile. For a pupil within the nation's highest SES quartile (SES 4), the chances of achieving above grade level is ten times that of a pupil from the nation's lowest SES quartile (SES 1). As can be seen below, the percent of pupils achieving above grade level increases exponentially from one SES quartile to the next (green). The percent of pupils achieving below the benchmark steadily decreases from each quartile to the next (blue). The percent of pupils achieving at grade level increases with each SES quartile, flattening between the two highest SES quartiles (red). This is largely due to the large increase in the above grade level percentage in the highest quartile (green).

Figure 9 Percent of pupil's below, at and above the grade level by SES quartile

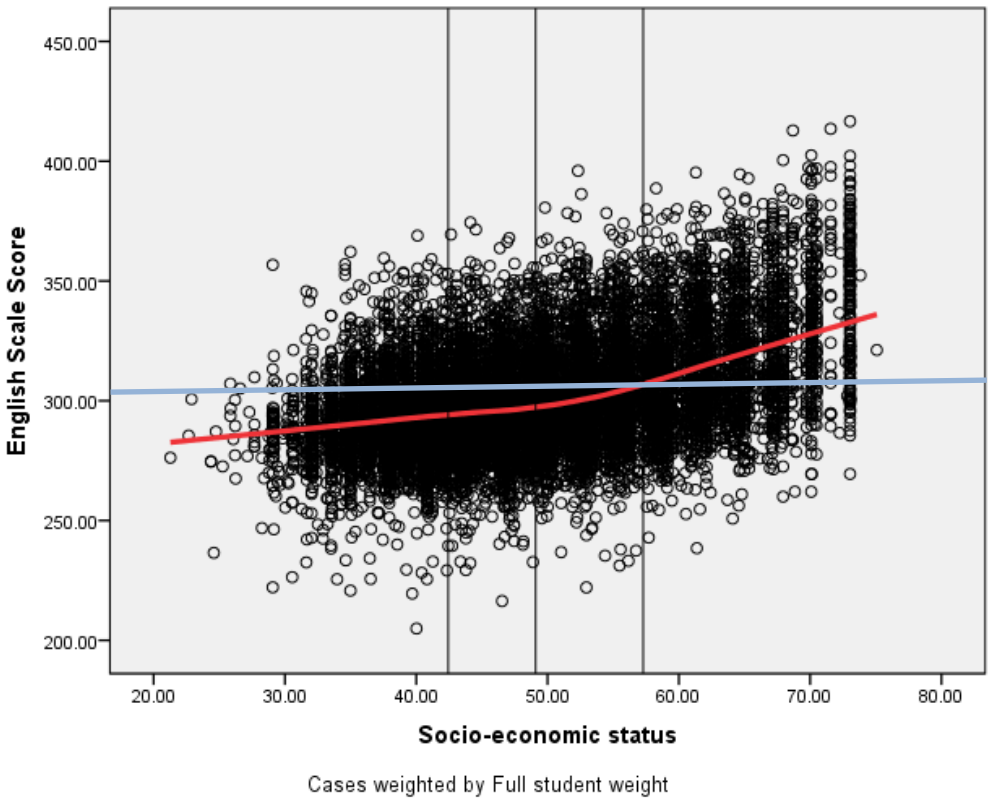


The analysis above provides a clear picture of the extent to which pupils are performing at, below or above grade level by SES quartile. This relies on categorising performance into three levels (below, at and above grade level) and categorising the whole population into four categories (SES quartiles). While providing a clear picture of the national situation, this approach can lose some important information. In order to assess, not only the effect of SES *between* SES quartiles, but the effect of SES on performance *within* each SES quartile another analysis was undertaken. Figure 10 represents English scale scores by socio-economic status (SES) along a continuum. The three vertical lines divide the SES continuum by the four population quartiles. Therefore the SES range 0 to 42.5 represent 25 per cent of the population with the lowest SES and the SES range 57 to 100 represent 25 per cent of the population with the highest SES. The red line represents the (curved) 'line of best fit' indicating the predicted English scale score, based on a pupil's SES. The (red) line of best fit indicates that as SES increases (y) so does the pupil's scale score in English (x). However, it is important to note that this line represents the *best prediction* of pupil performance based on SES, and the strength of the correlation determines if it is likely to be a good prediction (very likely).

As can be seen in the graph, the slope of the line is quite flat for pupils in the two lowest SES quartiles. The slope of the line increases in the highest SES quartile. The difference in predicted performance between those pupils at the lowest end of the SES Index and those pupils at the

highest end of the SES Index is quite large. The difference in predicted performance between pupils at the lowest end of the SES Index and the middle of the SES Index is quite small. This indicates that in Zimbabwe, SES has the greatest impact on performance for those pupils in the highest SES quartiles. The horizontal blue line represents the cut point for grade appropriate performance. As can be seen below, the predicted performance of pupils only reaches the grade appropriate level in the highest SES quartile. With this said, there is a large amount of variance between pupil performance across the SES continuum. There are still a large number of pupils (black dots) at the lower end of the SES Index performing above the predicted performance (red line) based on their SES and a large number of pupils at the higher end of the SES Index performing below the predicted performance.

Figure 10 English Scale Score by Socio-economic status



3.1.2 Language of instruction

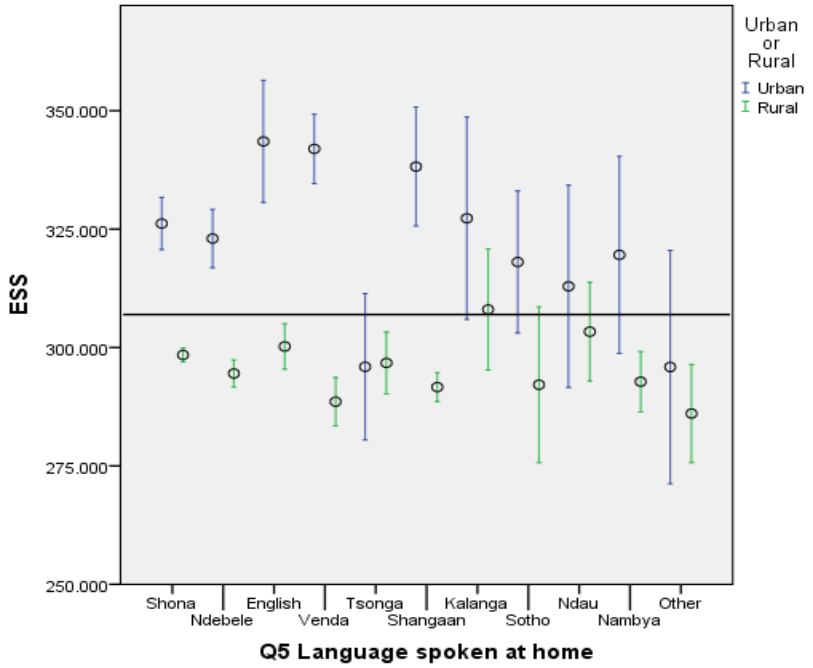
There was no evidence that pupils who are taught in languages other than the home language perform any better or worse on English and Mathematic tests. This was investigated by Province, yielding the same results.

3.1.3 Language spoken at home

Figure 11 represents the mean English scale scores for urban and rural pupils speaking each of the majority languages and minority languages at home. On average, urban pupils speaking English, Venda or Shangaan at home score higher on the English test than urban pupils speaking Ndebele and Shona at home. For pupils speaking Shona, Ndebele, English, Venda, Shangaan and Nambya, the location of the pupil appears to be a greater determining factor in performance than the

language spoken at home. However for Kalanga, Sotho and Ndau speaking pupils, though the mean score of urban pupils in urban locations is higher, the sample sizes are not large enough to know if this difference is real across the population. The patterns were similar for Mathematics achievement.

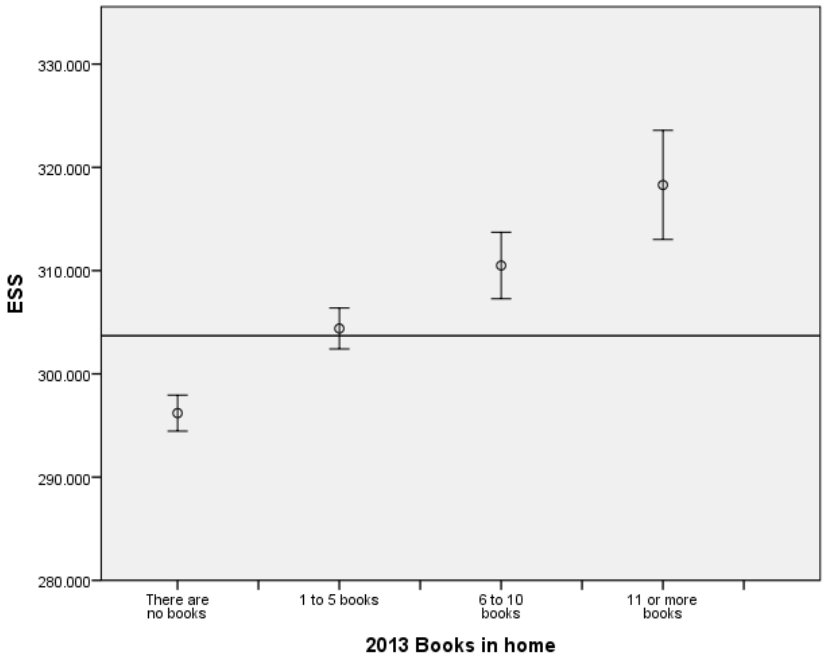
Figure 11 English scale score mean by language spoke at home and location



3.1.4 Number of books in the home

The number of books in a home tends to be correlated with pupil performance across countries and contexts. In Zimbabwe, there was a real and statistically significant difference between the average performance of pupils in each of the four categories: no books in the home, one to five books in the home, six to ten books in the home and more than eleven books in the home. The number of books in the home is a proxy for socioeconomic status, and therefore is largely immune to policy interventions except at the macro level. Interventions aimed at improving neonatal care, providing access to HIV medication and the provision of food security would be considered macro level interventions.

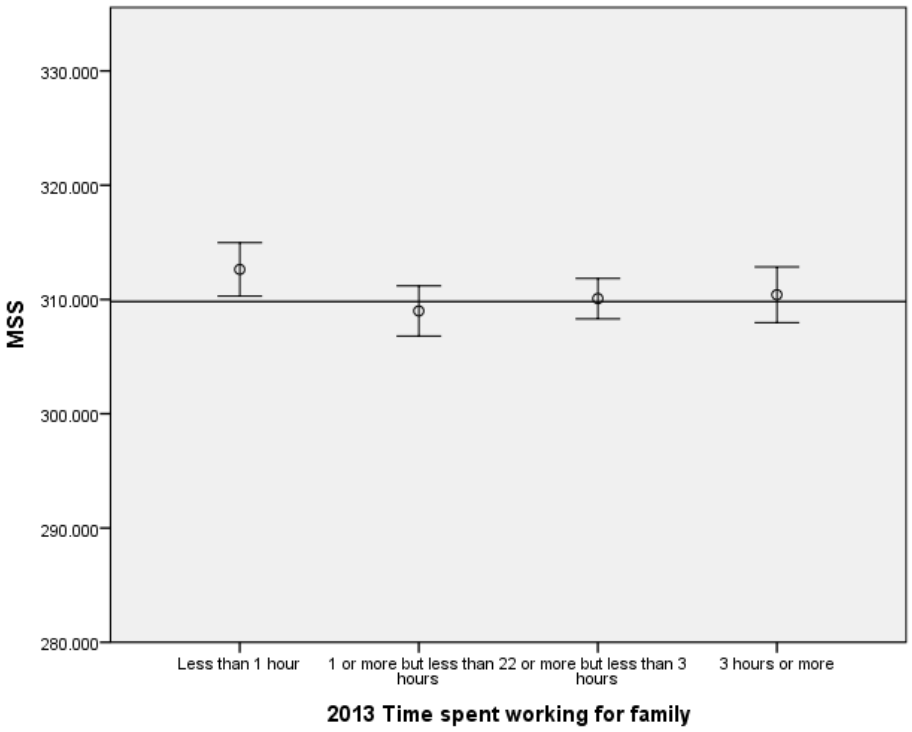
Figure 12 Mean English scale scores by number of books in the home



3.1.5 Time spent working for the family

In the 2013 cycle of ZELA, there was no evidence that the number of hours pupils spend working for family has any correlation with performance.

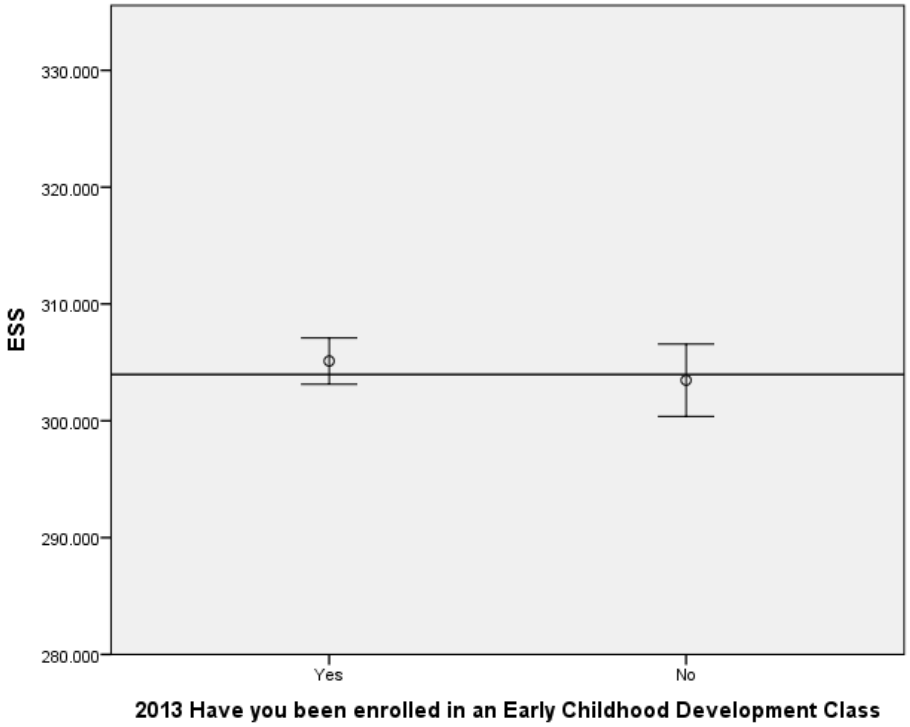
Figure 13 English scale score mean by number of hours spent working for family per day



3.1.6 Enrolment in an early Childhood Development Class

Unlike the findings at baseline, there was no evidence in 2013 that there was any real difference between the performance of pupils who were enrolment in an early childhood development class, compared to those pupils who were not.

Figure 14 Enrolment in an early Childhood Development Class



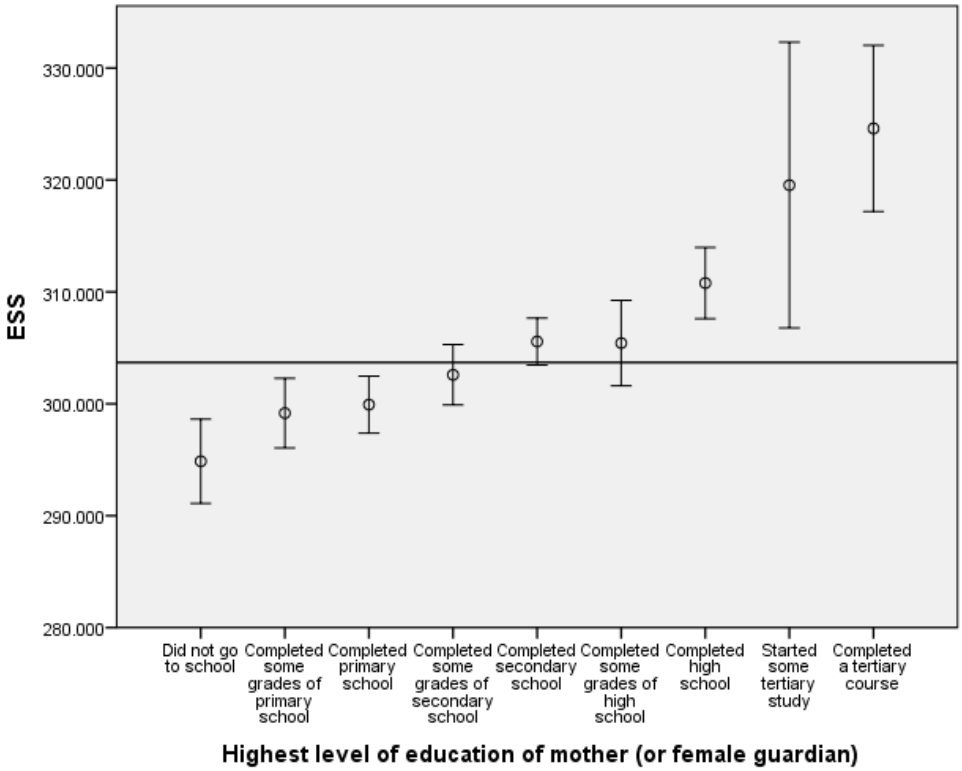
.

3.1.7 Parental education levels

Parent's highest level of education was strongly correlated with pupil performance. As can be seen in

Figure 15, as the mother's education level increases, so does pupil performance. Again this is in line with other international studies and is a proxy for socioeconomic status. Real differences in pupil performance appear to only be observable when the pupil's mother is reported to have completed the next level of education. For example, there is no evidence that the mean score of pupils whose mothers completed some grades of secondary school are any different to those pupils whose mothers finished primary school. This pattern can also be observed for the completion of some grades of high school and the completion of some tertiary education. This pattern was the same for father's education.

Figure 15 English scale score mean by highest level of education of mother

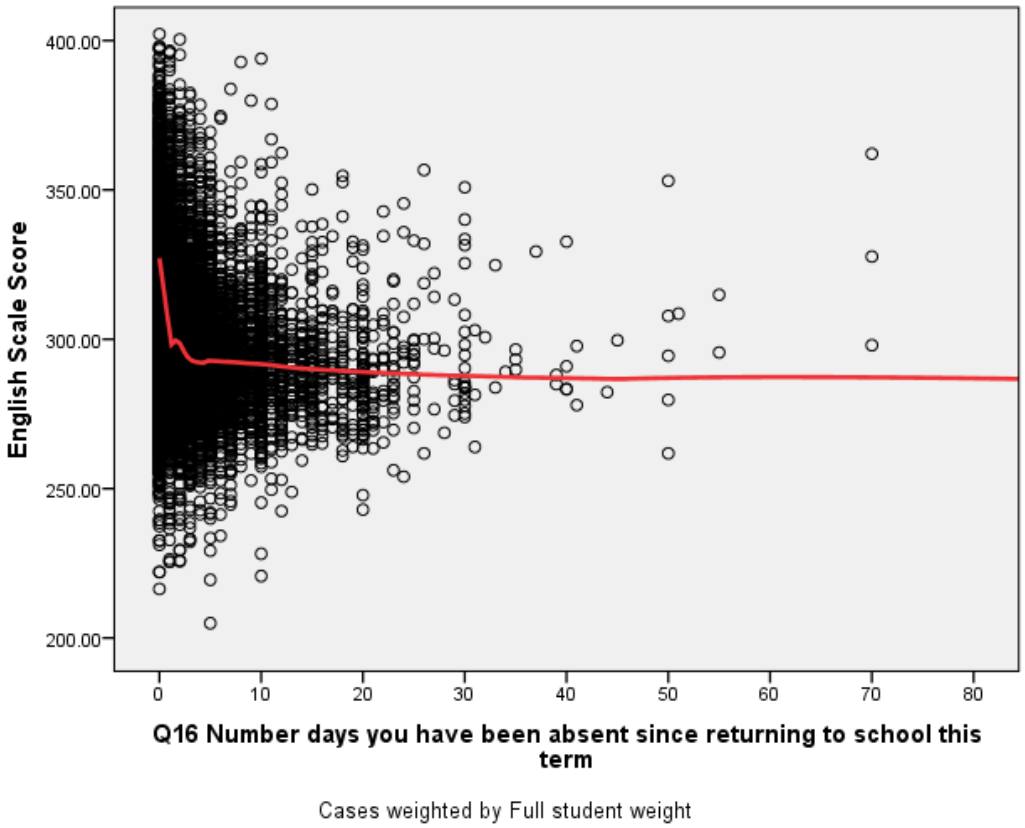


3.1.8 Pupil Absence

The pupils who were absent to care for ill family members at least one day in the previous term, achieved below the national average in English. The difference in English performance between those pupils who had not missed any days, and pupils who had missed one day or more the previous term was statistically significant.

Approximately a third of Grade 3 pupils in Zimbabwe reported that they had not been absent in the proceeding term and cumulatively, approximately half of the Grade 3 pupils in Zimbabwe were not absent or were only absent for one in the proceeding school term. The Figure below shows that the predicted performance of pupils missing up to five days of school in a term is significantly lower than pupils not missing any school. After five days of absence in the term, changes in the predicted performance of pupils are negligible.

Figure 16 English Scale Score by number of pupil days absent in the term



3.2 Teacher and teaching resources

An analysis of school level data related to the teacher and teaching resources available within schools was undertaken.

This section examines the relationship between test scores and various teacher and teaching resources. These include:

- Textbook use
- Pupil/textbook ratio
- Number of days teachers were absent
- The proportion of teachers with a teaching qualification
- The proportion of teachers who had attended professional development

On average, there was no evidence to suggest that pupils that are allowed to carry school textbooks home perform better than those who are not. Additionally, there were no or weak correlations between the number of activities the school uses the textbooks for (classroom instruction, library work and homework) and performance in English.

In 2013 there was a fairly moderate association between the pupil/textbook ratio and English performance, in urban schools. There were only very weak or no associations found between pupil achievement levels and:

- the use of textbooks for classroom instruction
- the use of textbooks for library work

There was a weak to moderate, statistically significant correlation between the pupil/textbook ratio in urban schools. This wasn't the case in rural schools. This may be due to the recent textbook distribution. Schools that have textbooks for pupils have not previously had access to textbooks, therefore the impact of increased access may not yet be observable.

There was a weak, yet statistically significant negative correlation between the UNICEF workbook ratio and socio-economic status in rural schools. Based on the findings related to socioeconomic status and the performance of pupils, this indicates that the UNICEF workbooks have been distributed to where they are most needed.

As can been seen in Table 8, there were weak or no correlations between:

- the number of days teachers were absent on official school days
- the proportion of teachers with a teaching qualification in the school (due largely to the high proportion of teachers who have teaching qualifications)[9]
- the proportion of teachers who had attended professional development.

Using Pearson's correlation

		English	Mathematics
Pupil/Textbook ratio	Urban	**.242****	**.156****
	Rural	.031	.009
Percentage of Teachers that are qualified	Urban	**-.176****	**-.088****
	Rural	**.188****	**0.169****
Number of School days lost	Urban	-0.121	-0.111
	Rural	0.052	0.005
Teacher absenteeism	Urban	-0.029	-0.072
	Rural	**-0.06****	-0.041

** Statistically significant

Table 8 Relationships between English and Mathematics Performance and Teacher and Teaching Resources

[9] This variable is further investigated through the Multi Level Model in Chapter Four.

3.3 School resources

In 2013 there was a fairly moderate negative correlation between pupil/toilet ratio and pupil performance in English in urban school, indicating that as the number of pupils sharing a toilet increases, performance decreases. However the correlation was not statistically significant. In urban schools, there was a weak, yet statistically significant correlation between performance and access to piped, tank or spring water at the school. In rural schools the correlation was moderate and statistically significant. Similarly, in rural schools access to electricity was weakly correlated with performance.

There were no or weak correlations between the following school variable and school resources:

- Pupil/Toilet ratio
- UNICEF textbooks
- Pupil/Permanent classroom ratio

		English	Mathematics
Pupil/Toilet ratio	Urban	-.2	-.120
	Rural	.034	.037
Pupil/Permanent classroom ratio	Urban	.002	.036
	Rural	-.016	-.058
Piped, tank or spring water	Urban	**.193****	**.158****
	Rural	**.270****	**.222****
Electricity (mains, generator or solar)	Urban	.055	**.057****
	Rural	**.166****	**.152****
Budget per pupil	Urban	-.004	-.033
	Rural	.040	.036

***** Statistically significant***

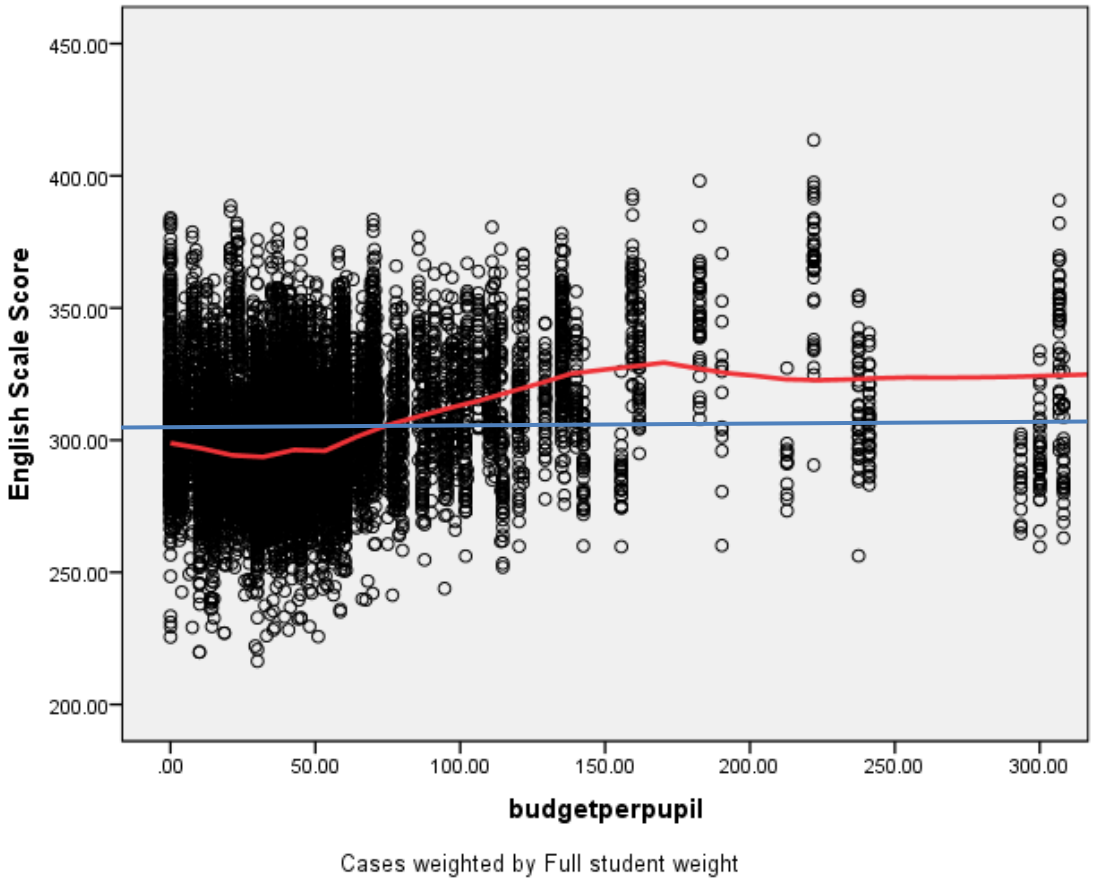
Table 9 School Infrastructure

In 2013 there were no or very weak associations between the school budget per pupil and achievement levels in English and mathematics achievement. Based on information from other studies a decision was made to further investigate the relationships between school budget and performance. An analysis of the average performance of pupils in schools by budget per pupil was carried out using an error graph. As found in the 2012 analysis, the graph showed that there is little, on average, difference between the performances of pupils based on funding. Using a regression analysis, the predicted performance of pupils based on the school budget per pupil was calculated by budget per pupil across the budget range of US$0 and US$320. The

analysis found the predicted performance of pupils only reaches the grade appropriate level once the per pupil school budget reaches US$75. However, it should be noted that the accuracy and reliability of such predictions depend on the strength of the correlation between the two variables. As can be seen in the graph, there remain many pupils achieving well above and well below the predicted performance line, across the budget spectrum. This means while the line represents the best prediction (the most likely performance of pupils based on school budget), it does not necessarily represent a good prediction (the very likely performance of pupils based on school budget).

There is little difference in the predicted performance of pupils at schools with a budget of less than US$50. Additionally there is little difference in the predicted performance of pupils at school with a budget of between US$175 and US$300 per pupil. However, there was a large difference in the predicted performance of pupils attending schools with a per pupil budget of between US$50 through to US$175. In other words, there appears to be two thresholds below and above which the level of budget allocations are less likely to result in impacts on pupil performance.

Figure 17 English Scale Score by budget per pupil



Chapter 4 To what extent can improvement in test performance be attributed to the Education Transition Fund?

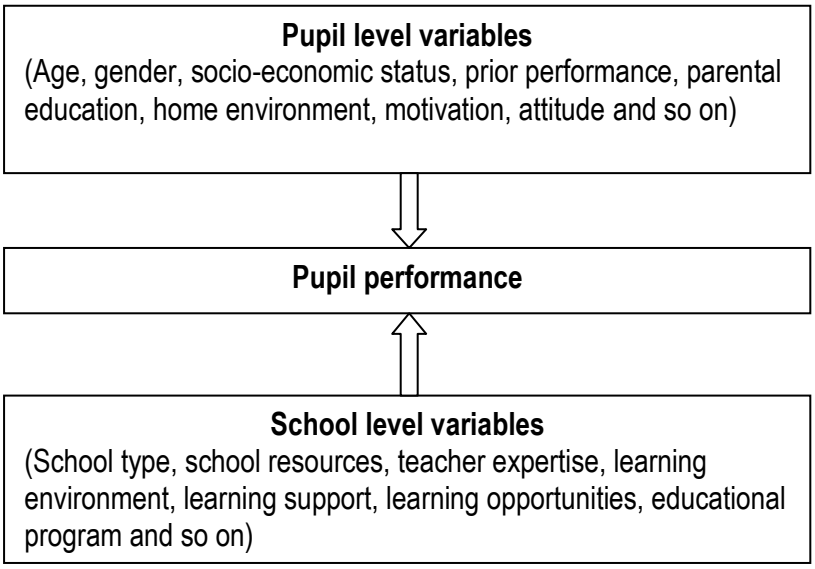
This chapter reports the third research question of the study; *To what extent can improvement in test performance be attributed to the Education Transition Fund?* As ZELA 2013 is a monitoring cycle, this question cannot be answered in its entirety. The ZELA study is correlational in its approach (examining the strength of associations) and due to the nature of the intervention there are no opportunities to monitor changes against a control group. Therefore, methodologies to answer the question of attribution must be flexible and dynamic throughout the course of the four year study.

4.1 Possible influences on pupil performance

As first step in the analysis, ZELA 2012 pupil performance data were linked to data from the pupil questionnaires and the school head questionnaires through school identification numbers that had been provided to schools. Of the 422 schools participating in ZELA 2012, some did not record an identification number and some recorded an identification number that was differed from the one in the pupil data. Thus in 2012 data from only 247 schools could be linked. Results reported in the 2012 ZELA baseline report were preliminary only, and designed to highlight how methodologies could be applied as part of the analysis in future cycles of the study.

Figure 18 depicts an example of how pupil performance might be influenced by variables at the pupil level and by variables at the school level.

Figure 18 Possible influences on pupil performance



This example aligns with the theoretical model described in Appendix One of this report (see input and output data).

The type of data collected has an hierarchical structure with pupils nested within schools. Because pupils within the same school share certain experiences, outcomes within schools might be more similar than outcomes between schools. Further, it is not only shared experiences within schools that are important. Pupils come from the same communities, and particular schools may select from

within those communities, thus schools are likely to be more homogeneous on input as well as process. Multilevel modelling offers the opportunity to formulate explicitly and to test hypotheses that pertain to each of the two levels of the hierarchy (individual pupil level and school level). The variance components can be partitioned among different levels (in this case two) so that there can be an answer to the question: *What proportion of the total variation in pupil outcomes is explained by variables at two different levels*? and further, *What characteristics of pupils and schools might explain this variation?*

4.2 Analysis

Analysis of the linked data was based on a two-level model: a within-school model at level 1 and a between-school model at level 2. The data were analysed using the multilevel modelling software HLM (Raudenbush, Bryk, & Congdon, 2004).[10]

Results from English and Mathematics tests were used for this preliminary analysis.

For each set of test data, three models were estimated:

1. An unconditional model without predictors (Model 0)
2. Model with pupil predictors added at level 1 (Model 1)
3. Model with pupil predictors at level 1 and school predictors added at level 2 (Model 2).

Pupil-level predictors used were:

- Age
- Gender
- Books at home
- TV
- Work for family
- Meals per day
- Early Childhood Development (ECD) class
- Socio-economic status
- Parental education
- Pupil absence to care for an ill family member
- Computer
- Calculator
- Religion

School-level predictors used were:

- School sector

[10] Raudenbush, S.W., Bryk, A.S, & Congdon, R. (2004). HLM 6 for Windows [Computer software]. Skokie, IL: Scientific Software International, Inc.

- Number of sessions
- Textbooks per pupil
- UNICEF distributed textbooks per pupil
- Percentage of qualified teachers
- Percentage of all teachers attending professional development (PD)
- Teacher absence
- Budget per pupil

The unconditional model enables partitioning of the variance of performance scores into the within-school component and the between-school component. The intra-class correlation measures the degree of similarity of pupils within schools and is given by the proportion of total variance that is attributed to the between-school component. A high value of the intra-class correlation indicates a high degree of homogeneity within schools.

Table 10 and

Table 11 display the parameter estimates for each of the three models for English and Mathematics, respectively.

Table 10 Multi-Level Model estimates for English Proficiency (EPV1)[11]

Parameter	Model 0	Model 1	Model 2
		Fixed Effects (Standardised)	
Level 1 (Pupil Level)	N=16,072	N=9,407	N=5,697
Age		-0.04 (0.01)*	-0.03 (0.01)*
Gender (Male=1, Female=2)		0.12 (0.01)*	0.13 (0.01)*
Books at home		0.11 (0.0.02)*	0.13 (0.02)*
TV			
Work for family			
Meals per day		0.03 (0.01)*	0.02 (0.02)
ECD		-0.02 (0.01)	-0.02 (0.01)
SES		0.16 (0.03)*	0.11 (0.05)*
Parental education		0.05 (0.02)*	0.06 (0.02)*
Absences for family care			
Computer			
Calculator		0.03 (0.02)*	0.05 (0.02)*
Religion - Traditional			
Religion - Apostolic			
Religion – Other Christian			
Religion - Jewish			
Religion - Muslim			
Religion - Other			
Level 2 (School Level)			
Government			
Private			
Church			0.11 (0.05)*
Sessions			0.21 (0.06)*
Textbooks per pupil			0.30 (0.12)*
UNICEF Textbooks			-0.34 (0.10)*
Teaching Qualifications			0.18 (0.07)*
Teacher PD			
Teacher Absences			
Budget per pupil			0.26 (0.07)*

[11] The multi-level model analysis was undertaken using each of the five plausible values for both English and Mathematics. There were limited significant effects across the five plausible values, but the effects were stable across all five PVs. See Appendix 3 for further information.

Variance Components Unstandardised)			
Within school	261.75	234.69	241.15
Between schools	263.43	189.19	120.09
Total	525.18	423.80	361.24
Intra-class correlation	0.50		
Proportion of Total Variance Explained		19.30%	31.22%

Note: Standard Errors in Parentheses- asterix indicates pr.<0.05

Table 11 Multi-Level Model estimates for Mathematics Proficiency (MPV1)[12]

Parameter	Model 0	Model 1	Model 2
	Fixed Effects (Standardised)		
Level 1 (Pupil Level)	N=16,072	N=9,407	N=5,697
Age			
Gender (Male=1, Female=2)		0.07 (0.01)*	0.07 (0.01)*
Books at home		0.13 (0.0.02)*	0.14 (0.02)*
TV		0.03 (0.01)*	0.02 (0.02)
Work for family			
Meals per day		0.07 (0.01)*	0.05 (0.02)*
ECD		-0.01 (0.01)	0.00 (0.01)
SES		0.14 (0.03)*	0.12 (0.04)*
Parental education		0.06 (0.02)*	0.05 (0.02)*
Absences for family care			
Computer			
Calculator			
Religion – Traditional (vs 'No Religion)			
Religion - Apostolic			
Religion – Other Christian		0.11 (0.04)*	0.09 (0.05)
Religion - Jewish			
Religion - Muslim			
Religion - Other		0.04 (0.02)*0	0.05 (0.03)
Level 2 (School Level)			
Government (vs 'Other)			
Private			
Church			
Sessions			0.23 (0.07)*
Textbooks per pupil			
UNICEF Textbooks			-0.23 (0.10)*
Teaching Qualifications			0.22 (0.07)*
Teacher PD			0.07 (0.07)
Teacher Absences			-0.04 (0.06)
Budget per pupil			0.11 (0.05)*

[12] The multi-level model analysis was undertaken using each of the five plausible values for both English and Mathematics. There were limited significant effects across the five plausible values, but the effects were stable across all five PVs. See Appendix 3 for further information.

Variance Components Unstandardised)			
Within school	310.39	272.12	280.89
Between schools	160.99	100.69	82.75
Total	471.38	372.81	364.64
Intra-class correlation	0.34		
Proportion of Total Variance Explained		20.91%	22.64%

Note: Standard Errors in Parentheses- asterix indicates pr.<0.05

4.2.1 Pupil –level variables that influence pupil performance

The pupil-level variables that appear to influence pupil performance (in order of effect size) in English are:

- Socio-economic status
- Gender
- Number of books at home
- Parental education
- Calculator as study material.
- Number of meals per day

The pupil-level variables that appear to influence pupil performance (in order of effect size) in Mathematics are:

- Socio-economic status
- Number of books at home
- Religion (Christian)
- Gender
- Number of meals per day
- Parental education
- Religion (Other)

4.2.2 School –level variables that influence pupil performance

The school-level variables that appear to influence pupil performance (in order of effect size) in Mathematics are:

- UNICEF Textbooks per pupil (negative effect)
- Textbooks per pupil
- Budget per pupil
- Number of sessions (Two sessions)
- Percentage of teachers with teaching qualifications
- School sector (Church school)

The school-level variables that appear to influence pupil performance (in order of effect size) in Mathematics are:

- UNICEF Textbooks per pupil (negative effect)
- Number of sessions (Two sessions)
- Percentage of teachers with teaching qualifications
- Budget per pupil

4.3 Equity

The analysis shows that for English, school effects are reasonably important, as compared to individual or family effects. Approximately the same level of variance in performance is explained by pupil level effects for English and Mathematics (19.3 and 20.9 percent respectively). However an additional 11.9 percent of variance (controlling for pupil level characteristics) is explained at the school level for English and only an additional 1.7 percent of the variance (controlling for pupil level characteristics) is explained at the school level for Mathematics. This could possibly be explained by the role of English in the early years in Zimbabwean schools and the use of English in society in general. Pupils from lower socio-economic conditions are less likely to be exposed to English.

Intra-class correlation (ICC) measures the proportion of the total variance in performance that can be explained by between-school variance. ICC is an indicator of education equity and lower levels of intra-class correlations represent higher levels of equity. This is because a low ICC indicates that schools are more similar, therefore making it less important which particular school a pupil attends. For ZELA, the ICC is .5 for English and .34 for Mathematics. This is quite high and indicates low levels of educational equity. The difference in target grade levels and domains between PISA Reading (15 year olds) and ZELA English ICC results limits the extent to which ICCs can be directly compared. However, for broad comparative purposes, Peru, Chile and Mexico have an ICC higher than Zimbabwe, Hong Kong and Japan have similar ICC levels and Korea, Thailand, USA, Australia, Canada and New Zealand have much lower ICCs (Cresswell, 2003). In countries where streaming takes place as part of the education sector structure, one would expect to see higher ICCs.

4.4 Targeting

Data collected in 2013 allows for analyses to inform the extent to which the ETF Program is effectively targeting the pupil's requiring intervention. The high negative relationship between pupil achievement and UNICEF distributed textbooks per pupil indicates that the ETF Program is effectively targeting pupils through the intervention. The indicator currently used to monitor pupil progress is the percentage of pupils achieving at a grade appropriate level. This logically signals the ETF Program intervention target population should be pupils achieving below the grade-appropriate level and schools and provinces with the greatest percentage of pupils achieving below the grade-appropriate level.

Analysis of the UNICEF distributed textbooks indicates that the textbook distribution has appropriately targeted these pupils, communities and provinces. Provinces with the lowest average pupil performance have a greater number of UNICEF distributed textbooks per pupil (on average). At the pupil level, the data also indicates that pupils achieving below the grade appropriate level have, on average, more UNICEF distributed textbooks than pupils achieving at or above the grade appropriate level.

Chapter Five Conclusion

Two hypotheses suggested by the analyses conducted in 2012 included:

1. An outcome of the intervention will be an increase in average achievement over the time of the intervention; and,
2. An outcome of the intervention will be a reduction in this proportion of school-level variance in achievement.

Chapter three of this report documents the increases in pupil performance in English and Mathematics across the majority of provinces in Zimbabwe. The patterns of achievement by province and gender are very similar for English and Mathematics, however the gap between urban and rural performance for English is greater than for Mathematics. The majority of provinces had a positive increase in performance in Mathematics and English and no provinces had a statistically significant decline in performance. Increases in Mathematics performance were greater than English performance, however the improvements made to the tests in 2013 may also have had an impact on the equating process (this can be determined in 2014).

Girls continue to outperform boys, with the gap remaining approximately the same size as in 2012 and Socio-economic status remained a strong predictor of success. There was no evidence that the performance of pupils in satellite schools had improved significantly.

The ETF Program has appropriately targeted pupils for textbook distribution, so based on the relatively low base some pupils may be starting from, combined with increasing exposure to reading materials, we would expect to see rapid advancements in pupil performance over the ETF Program cycle.

The school-level variance in performance was found to be relatively high, indicating that which school a pupil attends has a large effect on their performance. In line with the aims of the ETF Program, we would expect to see a reduction in the proportion of school level variance over the ETF Program cycle. The proportion of school-level variance in achievement will continue to be monitored and reported on over the 2014 ZELA monitoring cycle and will be used to determine the extent to which changes in pupil performance can be attributed to the ETF Program. Determining this attribution is one of the main aims of the wider ETF evaluation study.

5.1 Policy implications

One of the biggest predictors of pupil achievement is their ability. And so we would expect that a large proportion of variance will be explained at this, the pupil level. Pupil ability is, except from the most macro of perspectives (health, nutrition, ante-natal and post-natal care and so on), largely immune to policy changes. Pupils come to school with a more or less fixed amount of ability. Therefore, any variance that occurs at the school level is especially important because this can often be amenable to policy initiatives.

Based on the available evidence, the variables of particular interest in this study (from a policy perspective) include textbooks per pupil, budget per pupil and the percentage of teachers with a teaching qualification. These variables are open to policy intervention and have been found to have an effect on pupil performance.

Both the multilevel model (Chapter 4) and the analysis of the predicted performance of pupils based on school budget (Chapter 3); indicate that school budgets have a relatively significant effect on pupil performance. The analysis of predicted pupil performance by school budget per pupil suggests that the range of US$75 – US$175 per pupil is the most effective (and cost-effective) budget range to increase pupil performance.

The percentage of teachers with a teaching qualification was found to be a predictor of success for both English and Mathematics. This suggests that an effective strategy to improve pupil performance might include working with teachers colleges to increase the number of trained teachers and working with districts and provinces to support both unqualified teachers and schools with high proportions of unqualified teachers.

References

Cresswell, J. (2003) *The association between learning environment and student engagement and performance in reading: Results from the Pacific Rim in PISA 2000.* Paper presented at PISA International Conference.

Lin, C, Tzou, H. Hung, P and Huang, H. (2006). Comparisons of effect size of social and cultural status (ESCS) on literacy in various subject areas and different grades at school and student levels – Multilevel analysis of 2006 PISA data for Taiwan.

Ross, K.N., Pfukani, P., Nzomo, J., Makuwa, D., Machingaidze, J., Milner, G., Kulpoo, D., Postlewhaite, T. N., & Leite, S. (2000). *Translating educational assessment findings into educational policy and reform measures: Lessons from the SACMEQ initiative in Africa*. Paris: UNESCO.

UNESCO. (2003b). *Southern African Consortium for Monitoring Educational Quality (SACMEQ).* Harare: Author.

Matters, G. N. (2009). A problematic leap in the use of test data: From performance to inference. In J. J. Cumming and C. Wyatt-Smith (Eds.) *Educational Assessment in the 21st Century: Connecting theory and practice*. Dordrecht, The Netherlands: Springer International.

Appendices

Appendix 1: Theoretical framework

This chapter provides the theoretical framework for the research study and describes sources of data, how outcomes are to be measured, and the approach to the research.

The input–process–output model for data in the pupil learning environment is represented diagrammatically in Figure 19.[13] The entries in the boxes are illustrative. The framework is an adaptation of the 3P model of learning and teaching developed by Biggs (Biggs, 1999; Biggs & Moore, 1993), which portrays learning as an interactive system, identifying "three points of time at which learning-related factors are placed: presage, before learning takes place; process, during learning; and product, the outcome of learning" (Biggs, 1999). Biggs's model draws attention to two sets of presage factors: meta-contextual factors and those factors specific to the learner. In the adaptation of his model to datasets, the presage components are data about pupils, teachers, and school organisation and resourcing. This model is capable of generating predictions and associations that are relevant to this study and the policy questions most likely to be addressed by it. Reading from top to bottom, from input through process to output, the diagram portrays the storyline for an individual pupil or pupil cohort.

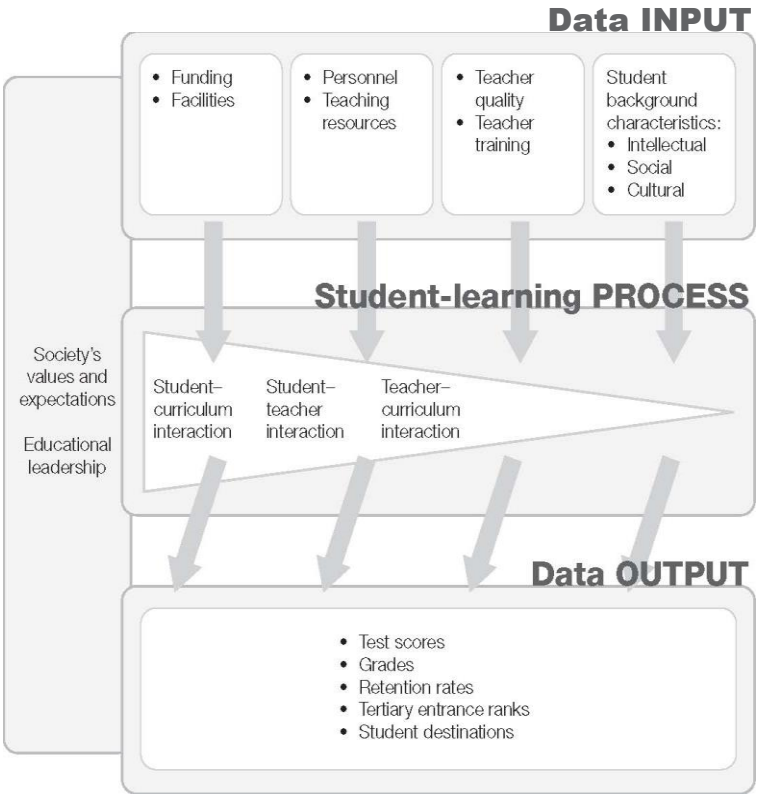


Figure 19 Input–process–output model for pupil data

A preliminary scan of reports[14] undertaken in early 2012 indicated that in Zimbabwe there was a severe shortage of teaching and learning materials and furniture in schools, most teachers needed

[13] Matters , G. N. (2006). *Using data to support learning in schools: students, teachers, systems* (Australian Education Review No. 49). Camberwell: ACER.

[14] *Education at a Glance 2009.* Zimbabwe Ministry of Education, Sport, Arts and Culture, Summary of Cost and Financing of the Education Sector in Zimbabwe. February 2010, *Education Statistics: Zimbabwe.* UNICEF Division of Policy and

professional development, and there was a significant proportion of orphans and vulnerable children in the population. Thus the study of the ETF program in Zimbabwe has gathered data about funding, facilities and resources, teacher quality and teacher training, and the pupils' backgrounds in both 2012 and 2013. Table 12 lists the data that were gathered as inputs for the model. Items appearing in black indicate inputs for the model in 2012 and 2013. Items appearing in blue indicate inputs gathered in 2013.

Table 12 Datasets – Input for Zimbabwe Study

Pupil level **(pupil background characteristics)**	**Teacher level** **(teacher quality and teacher training)**	**School level:** **Funding and facilities**
• Type of school attended • Age • Gender • Language spoken at home • No. of books in the home • Infrastructure accessed by the home (electricity and water) • No. of hours/day working for family/community • No. of meals eaten per day • Access to Early Childhood Development class • No. of days absent in Term 1 • Access to resources to study in the home • Socioeconomic status (parental education level) • Religion • Caregivers living in the home • No. of days absent in Term 1 due to caring for an ill family member	• No. of teachers • Qualifications of teachers • Gender • Teacher absentee rate • Professional development attendance • Qualifications of Head Teacher • Professional development attendance for teachers in the early years	• Province • School type • District • Language of instruction in the early years • Years of operation • Proximity to a large city • Pupil population – enrolled • Grade 3 pupil population – enrolled • Average class size • Minutes per lesson • Sessions per day • No. of days of closure of school operations • School infrastructure • Wash facilities • Orphans and vulnerable children (%) • Funding • Pupils with chairs (%) • Textbook supply • Textbook use • Pupils in fee arrears (%) • No. pupils with a disability • School days lost for pupils with caring responsibilities • Head Teacher professional development attendance (school management)

Practice, Statistics and Monitoring Section, 2008, Nyanguru and Peil, 1991. *Zimbabwe Since Independence: A People's Assessment*. African Affairs, 1991, 90, 607–620.

Data outputs are pupil scores on a test of language (two subtests) and mathematics (two subtests). Table 13 provides an overview of these outputs.

Table 13 Datasets – Output for Zimbabwe study

Subtest score: Operations	Subtest score: Application	Subtest score: Language	Subtest score: Comprehension
Score for Mathematics		Score for Language	

Appendix 2: Indicative standards

Indicative standards for three levels of achievement for mathematics

Pupils achieving at a grade-appropriate level can generally:

- Compare the magnitude of numbers below 100 and order objects from first to twentieth
- Recall simple multiplication facts when 2, 5 and 10 are factors
- Apply sets to solve multiplication problems and division problems using brackets
- Read time on the hour and on the half hour on a clock face
- Interpret pictorial representations of fractions
- Compare the area of plane shapes without the use of standard units
- Measure length in standard units
- Add two numbers in real-life context
- Count, add, compare and order numbers or objects in an unfamiliar context.

In addition, pupils achieving above a grade-appropriate level can generally:

- Count and write the number of objects in a given set
- Interpret the meaning of additional terms such as altogether, sum of n total
- Add two and three single- and double-digit numbers to give a total of up to 50
- Find the difference between any two numbers in the range 0 – 50
- Compare the length of various objects and the capacity of various containers without the use of standard units.

Pupils who are not yet achieving at the grade-appropriate level can generally:

- Find the product of two numbers by counting sets when the product is less than or equal to 100
- Recall basic division facts for multiples of 2, 5 and 10 up to 100
- Find a quotient by sharing equally when the dividend is less than or equal to 100
- Approximate numbers as nearer to 0, 10, 20 up to 100 Divide objects into halves and quarters
- Give the composition of Zimbabwean coins up to $1and calculate change for amounts not exceeding $1
- Read days of the week and place days of the week in their correct order
- Identify two-dimensional shapes and find the perimeter of these shapes using non-standard units
- Compare the mass of common objects and the rate of performed tasks without the use of standard units

- Interpret the meaning of a mathematical story, translate words into numerical statements, and solve the problem posed.

Indicative Standards for three levels of achievement for English

Pupils operating at the grade-appropriate level in English generally identify and name familiar objects in English, spell simple words correctly, and express in words the number of objects between 0 and 10. Pupils also know how to use simple prepositions with or without the help of a diagram, verb tenses (simple present, simple past, and simple future), possessive pronouns, and superlative forms of simple adjectives. They can use simple adverbs using the indefinite article (a/an), and use simple adverbs. Pupils locate detail in a text and find directly stated information. They connect ideas and make inferences.

Pupils achieving above a grade-appropriate level can generally identify and name familiar objects in English, spell simple words correctly, and express in words the number of objects between 0 and 10. Pupils also know how to use simple prepositions with or without the help of a diagram, verb tenses (simple present, simple past, and simple future), possessive pronouns, and superlative forms of simple adjectives. They can use simple adverbs using the indefinite article (a/an), and use simple adverbs. Pupils locate detail in a text and find directly stated information. They connect ideas and make inferences.

Pupils operating below the grade-appropriate level generally identify and name familiar objects in English, spell simple words correctly, and express in words the number of objects between 0 and 10.

Appendix 3: Significant effects in the MLM across PVs

Table 14 **Significant Effects in the Full model for EPV1 to EPV5 - Effect Significant at pr. <0.05**

	EPV1	EPV2	EPV3	EPV4	EPV5
Age	X	X	X		
Gender (Male=1, Female=2)	X	X	X	X	X
Books at home	X	X	X	X	X
TV					
Work for family					
Meals per day		X			X
ECD					
SES	X		X	X	
Parental education	X	X	X	X	X
Absences for family care					
Computer					
Calculator	X	X	X	X	X
Religion - Traditional					
Religion - Apostolic					
Religion – Other Christian		X			X
Religion - Jewish				X (-ve)	
Religion - Muslim				X (-ve)	
Religion - Other					
Level 2 (School Level)					
Government					
Private					
Church	X		X		
Sessions	X	X	X	X	X
Textbooks per pupil	X	X	X	X	X
UNICEF Textbooks	X (-ve)	X (-ve)	X (-ve)	X (-ve)	X (-ve)
Teaching Qualifications	X	X	X	X	X
Teacher PD					
Teacher Absences					
Budget per pupil	X	X	X	X	X (pr.=0.050)
Intra-class Correlation	0.50	0.50	0.51	0.49	0.50

Table 15 **Significant Effects in the Full model for MPV1 to MPV5 - Effect Significant at pr. <0.05**

	MPV1	MPV2	MPV3	MPV4	MPV5
Age					
Gender (Male=1, Female=2)	X	X	X	X	X
Books at home	X	X	X	X	X
TV				X	
Work for family					
Meals per day	X	X	X	X	X
ECD					
SES	X	X	X	X	
Parental education	X	X	X	X	X
Absences for family care					
Computer					
Calculator					
Religion – Traditional					
Religion - Apostolic					
Religion – Other Christian	X	X			X
Religion - Jewish					
Religion - Muslim					
Religion - Other	X	X	X	X	X
Level 2 (School Level)					
Government					
Private					
Church	X				
Sessions	X	X	X	X	X
Textbooks per pupil					
UNICEF Textbooks	X (-ve)	X(-ve)	X(-ve)		X(-ve)
Teaching Qualifications	X	X	X	X	X
Teacher PD					
Teacher Absences					
Budget per pupil	X		X		
Intra-class Correlation	0.34	0.35	0.34	0.34	0.34