

**siyaJabula  
siyaKhula**  
educates the nation



## kaMhinga Literacy Project



## Hope for rebuilding language foundations

Human Sciences Research Council Evaluation Report:  
siyaJabula siyaKhula's Learner Regeneration Project  
Mhinga, Limpopo  
Evidence after Year 1 (2013)

Prinsloo, CH & Rogers, S

May 2014



J.P.Morgan

# Table of Contents

<b>List of Tables, Figures and Annexures</b>	<b>2</b>
<b>Foreword</b>	<b>5</b>
<b>Executive summary</b>	<b>8</b>
<b>Introduction: The origins of the language intervention</b>	<b>11</b>
<b>1. Background</b>	<b>11</b>
1.1 Ethical clearance of the evaluation project	11
1.2 Literature	12
1.3 The evaluation brief	12
<b>2. Methodology and design</b>	<b>14</b>
2.1 Evaluation purposes / objectives	14
2.2 Approach to the study	15
2.3 Participants / sample	15
2.4 Instruments	16
2.4.1 Background questionnaires	16
2.4.2 Learner assessments	16
2.5 Evaluation procedures	17
<b>3. Evaluation findings</b>	<b>18</b>
3.1 Contextual background and its equivalence between project and control schools	18
3.1.1 School equivalence	19
3.1.2 Teacher and classroom equivalence	20
3.1.3 Home and learner contextual background equivalence	23
3.2 Impact of interventions	27
3.2.1 Grade 1	33
3.2.2 Grade 4	38
3.2.3 Grade 7	44
3.3 Comparing learner achievement between monitored and not-monitored test administrations	49
3.3.1 At Grade 1 level	53
3.3.2 At Grade 4 level	54
3.3.3 At Grade 7 level	55
3.4 Uptake factors and systemic issues influencing rolling out and scaling up the intervention	56
<b>4. Discussion</b>	<b>58</b>
4.1 Conclusions	58
4.2 Implications and recommendations	60
<b>Endnotes</b>	<b>61</b>

## List of Tables, Figures and Annexures

Table 1: Capacity components favouring project and control schools respectively	19
Table 2: Number of learners and classes on average by grade for project and control schools	20
Table 3: Existence, sufficiency and quality of school facilities for project and control schools respectively	20
Table 4: Capacity comparison for teachers from project and control schools	21
Table 5: Existence, sufficiency and quality of classroom infrastructure and facilities for project and control schools	22
Table 6: Availability and quality of materials used in classrooms for project and control schools	23
Table 7: Percentage of parents / caregivers indicating that learner homes have the listed items	25
Table 8: Hours of normal and intervention tuition received by school type and grade during the first semester	28
Table 9: Standard format of one-hour intervention sessions to Grade 1 learners in the first semester	29
Table 10: Standard format of one-hour intervention sessions to Grade 4 and Grade 7 learners in the first semester	29
Table 11: Hours of normal and intervention tuition received by school type and grade during the second semester	30
Table 12: Standard format of one-hour intervention sessions to Grade 1 learners during the second semester	30
Table 13: Standard format of one-hour intervention sessions to Grade 4 and Grade 7 learners during the second semester	30
Table 14: Comparison of learner contact time between control and project schools	32
Table 15: Mean percentage achievement, percentage point change, difference-in-difference and percentage change figures for Grade 1 learners by school type	35
Table 16: Percentage of learners in project and control schools achieving a mark of at least 50% in the four Grade 1 sub-tests during testing in February, June and November 2013	37
Table 17: Mean percentage achievement (age estimate for Schonell test), percentage point change, difference-in-difference and percentage change figures for Grade 4 learners by school type	39

Table 18: Percentage of learners in project and control schools achieving a reading age of 8 or higher or at least 50% in the seven Grade 4 sub-tests during testing in February, June and November 2013	44
Table 19: Mean percentage achievement (age estimate for Schonell test), percentage point change, difference-in-difference and percentage change figures for Grade 7 learners by school type	45
Table 20: Percentage of learners in project and control schools achieving a reading age of 11 or higher or at least 50% in the three Grade 7 literacy sub-tests and its total score during testing in February, June and November 2013	48
Table 21: Monitoring coverage at the baseline and after the first semester and first year	50
Figure 1: Simplified schematic view of quasi-experimental impact evaluation design	15
Figure 2: Illustrative schematic view of the result of difference-in-difference analysis for one cohort	15
Figure 3: Grade 1 mean learner scores (%s) on the EGRA Letter Sound sub-test over time by school type	35
Figure 4: Grade 1 mean learner scores (%s) on the EGRA Word Recognition sub-test over time by school type	36
Figure 5: Grade 1 mean learner scores (%s) on the EGRA Non-Word Decoding sub-test over time by school type	36
Figure 6: Grade 1 mean learner scores (%s) on the Oral Pictorial test over time by school type	37
Figure 7: Grade 4 learner mean scores (reading age) on the Schonell Reading Age test over time by school type	39
Figure 8: Grade 4 mean learner scores (%s) on the EGRA Letter Sound sub-test over time by school type	40
Figure 9: Grade 4 learner mean scores (%s) on the EGRA Word Recognition sub-test over time by school type	40
Figure 10: Grade 4 learner mean scores (%s) on the EGRA Non-Word Decoding sub-test over time by school type	41
Figure 11: Grade 4 learner mean scores (%s) on the Literacy 1 (MCQ) sub-test over time by school type	41
Figure 12: Grade 4 learner mean scores (%s) on the Literacy 2 (Sentence Ordering) sub-test over time by school type	42
Figure 13: Grade 4 learner mean scores (%s) on the Literacy 4 (Open-ended items) sub-test over time by school type	42

Figure 14: Grade 4 learner mean scores (%s) on the overall (aggregated) Literacy test over time by school type	43
Figure 15: Grade 7 mean learner scores (age) on the Schonell Reading Age test over time by school type	46
Figure 16: Grade 7 mean learner scores (%s) on the Literacy 1 sub-test (MCQ) over time by school type	46
Figure 17: Grade 7 mean learner scores (%s) on the Literacy 3 sub-test (Paragraph Writing) over time by school type	47
Figure 18: Grade 7 mean learner scores (%) on the Literacy 4 sub-test (open-ended items) over time by school type	47
Figure 19: Grade 7 mean learner scores (%s) on the total Literacy test over time by school type	48
Annexure 1: Comparative attrition (balance) statistics for Grade 1 learners: Paired and unpaired sub-groups	63
Annexure 2: Comparative attrition (balance) statistics for Grade 4 learners: Paired and unpaired sub-groups	64
Annexure 3: Comparative attrition (balance) statistics for Grade 7 learners: Paired and unpaired sub-groups	65
Annexure 4: Comparative monitoring statistics for Grade 1 learners: Monitored and not-monitored scores and gains for the project and control groups	66
Annexure 5a: Comparative monitoring statistics for Grade 4 learners: Monitored and not-monitored scores and gains for the project and control groups	67
Annexure 5b: Comparative monitoring statistics for Grade 4 learners: Monitored and not-monitored scores and gains for the project and control groups	68
Annexure 6: Comparative monitoring statistics for Grade 7 learners: Monitored and not-monitored scores and gains for the project and control groups	69

## Foreword

Just after the middle of 2013, a provisional evaluation report of the impact of siyaJabula siyaKhula's (sJsK's) learner regeneration intervention was produced six months into the implementation. The intervention was aimed at repairing critical gaps in learners' literacy and language foundations, especially those relating to decoding skills and understanding letter-sound relations. It was also aimed at automating these proficiencies which are assumed to be critical for reading fluency, comprehension and cognitive development.

With work-in-progress status, this first-semester evaluation report had readers from the funders and intervention agency as its main targets. In relation to the funders, being the United States Agency for International Development (USAID) and ELMA Foundation, it fulfilled reporting and funding obligations. For sJsK the information served to take stock of progress and informed mid-year adjustments to the interventions, as required. The evaluation findings reported were considered to be provisional in anticipation of completing a full year of intervention rollout. Setting up the initial structures and procedures related to and in preparation of the intervention took much effort and time. This left a rather short intervention implementation period for the first semester of 2013. A full year of intervention has now taken place. Besides the two immediate stakeholders, that is, the funding and implementation organisations, we have confidence now that an official report can be distributed to a wider audience. It is underpinned by more stable and longer-term data gathering. The recipients primarily include the beneficiaries of the intervention, but also a wider audience of officials from relevant government departments and school authorities, as well as academics. The beneficiaries are understood in the widest possible sense starting with the schools, teachers, learners and parents where the intervention was rolled out to a first cohort in 2013. Everyone deserves to know how the intervention is doing. Within the Department of Basic Education, those circuit, district and provincial authorities involved in the implementation, and its evaluation through the participation of control schools, comprise a key audience.

Because this report is the first formal or "official" one, relevant background and other information from the provisional Semester 1 report are retained in this version. Substantive integration of literature, context and findings about any observed learner achievement gains, including further theorising about the mechanisms at work, are reserved for later in the 2015 versions of the evaluation report, if not the final report only. One key reason for holding back on intensive evaluation is that scaling up of the programme to a second local language, Tshivenda, in addition to Xitsonga, and pursuing the first cohort of Grade 1 and 4 learners into their next grades, Grades 2 and 5, only commenced earlier in 2014. The findings from this second phase are therefore still forthcoming. Publication in academic journals is also reserved for subsequent outputs.

We next want to emphasise a few important points in relation to the contents and interpretation of the findings in this Year-1 impact report.

- A dramatic increase in learner achievement off a low baseline was expected, based on two factors. First, supporting reports and evidence accompanying sJsK's funding proposal and application documented how the learner regeneration intervention was "partly proven and promising". A second expectation was that learners from disadvantaged backgrounds and contexts usually respond very strongly initially to any substantive intervention.
- In line with what is sometimes known as "explosive learning"<sup>1</sup>, learners who suddenly grasp key foundational concepts have acquired a new toolkit that can be used without external mediation, a

---

<sup>1</sup> A literature is emerging on concepts such as accelerated or rapid language learning, often underpinned by cognitive approaches to language learning. Without citing specific references formally, as the programme implementers intend to publish on this approach in due time in peer-reviewed academic journals in order to explain and motivate their "theory of change" in depth, some indications of this emerging work are provided nevertheless. Skehan (Centre for Applied Linguistic Research; "A cognitive approach to language learning"; 1998) and Krashen and colleagues (Asher and Larsen-Freeman), through Oxford University Press, for instance, were some of the early proponents towards the turn of the millennium. In addition, the concept of "superlearning" is often related to the original work of the

modality to which they would have been exposed to extensively before. They therefore initially exceed the expected impact brought about only by mediated intervention, and add their own value, so to speak. Such a rapid period of growth was expected in the initial stages of this intervention, which would taper off over subsequent years. This does not detract from dramatic Year-1 growth. Sustaining improvement over longer periods is a dimension that interventions have to address and evaluation has to cover in the medium to longer term.

- Related to the previous point, the intervention has deliberately focused so far on specific foundational concepts and their mastery. These include components such as the alphabetic principle, letter sounds (phoneme-grapheme interactions), decoding, and the like. Giving focused attention to these, regular repetition, and mastery of minimum levels of fluency in processing, should create observable and strong early effects.
- In line with the previous point, learner assessment and the impact evaluation in this project have so far also focused on these early gains. In addition, the assessment and evaluation activities were extended at most thus far to components such as comprehension and some mastery of written language.
- On the basis of learners' ever expanding vocabulary and improving de/encoding skills, the next proficiency that should develop would be fluency of processing. This would enable learners to process meaningful pieces of new knowledge and learn new concepts by linking it to their existing levels of knowledge and understanding, within the short processing times that the so-called "working memory" affords a learner's brain. Gradually expanding conceptual knowledge, comprehension and broader academic proficiency will follow. Because this essentially is a hierarchical and cumulative process, gains in relation to generally improved school marks will follow incrementally after some lag time.
- In view of all the above, it is anticipated that more sophisticated across-the-curriculum learning and achievement effects will show only after two or more years, when consideration will be given to expand how academic achievement is measured. This is anticipated to include learners' marks from past and current Annual National Assessment. One advantage of conducting difference-in-difference analyses is that any teaching-to-the-test or coaching effects can be controlled for. As long as there is a common instrument at any given moment, say at the baseline, mid-term or summative point, for both project and control learners, comparisons remain valid. This remains true should one decide to select different end-line instruments at some point in the evaluation.

A final few brief comments now follow pertaining to the evaluation methodology and instruments. The Schonell Reading Age and Early Grade Reading Assessment (EGRA) are both internationally benchmarked. That means that their development ensures the appropriateness of their items and critical requirements for standardised implementation and scoring. Along with these, sJsk developed grade- and context-appropriate reading passages and comprehension items. These items were spread across different modalities, including multiple-choice items, sentence ordering, and short sentence and paragraph writing. As a result, content and construct validity and reliability were considered sufficiently covered. Funding and other capacity constraints, but in particular sJsk's ongoing need for information about learner achievement at multiple points to enable sJsk continuously to tailor learning materials and intervention implementation to the needs of individual learners, or small sub-groups of them, determined another specific choice related to the evaluation methodology. This entailed that the administration of assessments was retained as a key sJsk task. The Human Sciences Research Council's (HSRC's) evaluation team received these assessment data

---

Bulgarian educator and psychologist, Georgi Lozanov, and said to underpin Suggestopedia as approach. In the same breath names like that of Bobbi de Porter from the Balkan Business School are mentioned along with the notion of "accelerated encoding of long-term memory" ([www.demystifying.com/about-teaching-method.html](http://www.demystifying.com/about-teaching-method.html); accessed 14 Mar 2014). Michael L Anderson also wrote on "super-normal returns" when intervention programmes are delivered early (Journal of the American Statistical Association, 103(484), 1481-1495; 2008).

also for purposes of independent impact evaluation. This required an adjustment to the quality assurance coverage that is normally adhered to in determining assessment (evaluation) data reliability and validity. A decision was made to increase the customary quality assurance coverage from around 10% to over about 25% to 33%. Comparison of project- and control-school learner gains during evaluation factored in monitoring status at sites.

With regard to memory effects and coaching (teaching to the test), specific safeguards were built into either test selection or administration procedures and frequencies. The following issues are relevant:

- The Reading Age assessment (Schonell) used by sJsK and in the evaluation is a simple, but effective, and robust tool. For one, because of how it is administered and scored, it carries over no learning or memory effects. The reason is that administration of this test is terminated after a learner makes five mistakes, ensuring that next time around, in assessments repeated every semester, learners always get to new or hitherto unseen levels of the instrument. With each assessment learners will encounter new and more difficult words, assuming that learning has taken place.
- For EGRA, learners are expected to improve rapidly on their initial scores. In terms of the pre-agreed methodology of double-difference analysis, even if there would be unobservable influences on learners' scores the second time round, such effects would be consistent across school type, that is, the same for project- and control-school learners. This design feature even allows the evaluation team to replace any initial assessment tool at any stage with another should there be reason to believe any initial instrument is no longer yielding useful data.
- Teachers, facilitators and learners are given no access to test materials. They therefore at most have fleeting glimpses of their contents only to see the instruments again at least six months later.

Selected explanations and confirmations have been provided above. In addition, information inside the report explains the methodological approach to this evaluation more fully.

Last, it is confirmed that two external readers made extensive comments to a previous version of this manuscript. These have all been addressed. Another set of revisions then followed internal HSRC review by the appropriate line-function manager. The report is hereby released in the belief that it adds a substantive first evaluation of the intervention, and new knowledge about what has been possible in terms of approach and effects.

---

Dr V Reddy: Executive Director  
Education and Skills Development (ESD) Research Programme  
Human Sciences Research Council



---

*This report is made possible by the generous support of the American people through the United States Agency for International Development (USAID). The contents are the responsibility of the Human Sciences Research Council (HSRC) and do not necessarily reflect the views of USAID or the United States Government.*

---

## **Executive Summary**

### **Introduction and background**

In June 2011 siyaJabula siyaKhula (sJsK) applied for a grant from the United States Agency for International Development (USAID). The subsequent grant award, from the USAID's School Capacity and Innovation Program (SCIP), was augmented by the Elma Foundation and enabled sJsK to expand and test a three-year language development intervention in primary schools. The grant also provided for an impact evaluation of the intervention by the Human Sciences Research Council (HSRC).

### **Key purpose and objectives of the evaluation**

The evaluation focused on the cost-effective impact of a learner regeneration intervention implemented in the Mhinga Villages and surrounding areas and the potential for its refinement and expansion from local operational levels to broad-based provincial and national scale. In order to achieve this, the evaluation work included establishing contextual and learner achievement baselines for the participating project and control schools, comparing and confirming sufficient equivalence between these schools' background situations and comparing the relative achievement gains among learners from the two groups.

### **Evaluation design and methodology**

The evaluation follows an experimental- and control-group design. At the outset, it determined the contextual conditions at schools, in classrooms and at home, as well as learner achievement levels. All learners in Grades 1, 4 and 7 in 11 project schools in 2013 comprised the first of three cohorts of the intervention population. Their anticipated achievement gains are compared with those of all learners from five matched control schools. The evaluation data were collected by means of self-report background questionnaires completed by school principals, teachers and parents/caregivers, as well as a range of language assessment instruments administered among/to the learners. Difference-in-difference analysis was conducted to establish relative achievement gains among learners from project and control schools.

### **Findings pertaining to equivalence of project and control school**

The capacity and responsibilities among school staff and infrastructure, facilities and conditions at schools were found to be largely equivalent between the project and control groups. All schools are located in a deeply rural environment, characterised by socio-economic circumstances of unemployment and poverty. Teacher capacity and responsibilities were also greatly equivalent between the two groups. So were classroom infrastructure, facilities and materials. Differences that were adjudged to be practically meaningful were limited in number. These were often not directly language-related. Parental/caregiver or home characteristics and profiles, socio-economic status, learner access to schools and their exposure to language materials and support were quite homogenous. Communities are poor and predominantly Xitsonga-speaking. One would therefore with high confidence be able to attribute differing achievement gain trajectories between learners from project and control schools to the interventions in view of the large equivalence between the project and control groups.

## Findings pertaining to intervention impact after one year

After a full year of programme implementation, including unavoidable delays during the start-up period, significantly greater benefits to project-school learners were evident.

Grade 1 learners' proficiency in all four sub-tests (Letter Sound, Word Recognition, Non-Word Decoding and Oral Pictorial) showed much stronger gains among learners from project schools compared to those from control schools. However, it is in relation to the more technical decoding proficiencies (Letter Sound and Non-Word Decoding) that learners from project schools improved most compared to control-school learners. Where low baselines (close to zero) do not render growth calculations totally meaningless, the findings reveal improvements of around 600% (with relatively low baselines still explaining these high figures partly) on average for learners in project schools compared to improvements of between 150% and 340% for learners in the control group on the Letter Sound and Oral Pictorial (Vocabulary) sub-tests respectively. Whereas beyond 70%, but mostly close to or exceeding 90%, of learners from project schools by the end of the year achieved a mark of at least 50% on these individual sub-tests, the corresponding percentages for learners from control schools remained below 20%, with the exception being that almost 60% of them achieved that level in the case of the Oral Pictorial (Vocabulary) sub-test.<sup>2</sup>

Grade 4 learners from project schools showed significantly higher gains compared to learners from control schools in all but one (Literacy Multiple Choice Questions - MCQ) of the eight available test scores. These comprise Reading Age (reading fluency and accuracy), a number of other reading skills (Letter Sound, Word Recognition and Non-Word Decoding skills) and Comprehension (sentence ordering and open-ended responses). These findings represent improvements ranging from almost 30% to as high as 268% and once over 400% for learners from project schools compared to changes from just more than 10% to almost 150% for learners from the control group. Reading Age gains for learners in project schools were slightly more than double (1,7 years) that of learners in control schools (0,8 years). Only 7% of control-school learners were at their age-appropriate reading age compared to 26% of the project-school learners. More than 90% of the learners in project schools achieved scores of 50% or more after a year in the three EGRA reading proficiencies (Letter Sound, Word Recognition, Non-Word Decoding). In relation to the latter two proficiencies, the comparable figures at control schools were at most 75%. The figure for the former proficiency of the three was 10%. In relation to receptive comprehension (multiple-choice and sentence ordering items), almost 15 percentage-points more project-school learners compared to control-school learners achieved 50% or more on these two sub-tests (32% and 47% respectively compared to 19% and 31%). Proficiencies within the direct aim of the intervention programme, i.e., phonemic and decoding skills, benefitted substantively, while improvements in derived competencies, such as comprehension, would most likely follow later after a not-unexpected time lag.

Grade 7 proficiency gains, as expected, given the extent of and duration over which conceptual gaps had developed, took a slower course. Nevertheless, gain score differences in all the test components were rather large between project- and control-school learners in favour of the former group. The findings pertaining to the comprehension or Literacy test scores represent improvements ranging between 80% and 460% for learners in project schools compared to changes from 30% to 70% for learners in the control group. Reading Age gains for learners in project schools were more than double (1,2 years) that for learners in control schools (0,5 years). Only 2% of control-school learners were at their age-appropriate reading age compared to 6% of the project-school learners. It should be noted that these learners have to bridge a gap of almost four years of work. In relation to the aggregated literacy score, 35% of project-school learners achieved a score of at least 50% compared to 11% of the control-school learners. The corresponding figures pertaining to the multiple-choice, paragraph writing and other written responses sub-tests were 74%, 60% and 21% respectively for project-school learners compared to the figures of 52%, 32% and 5% for control-school learners.

---

<sup>2</sup> Qualitative interpretation is favoured above statistically significant quantitative differences because of design and sampling features (learners clustering in a small number of homogenous rural schools). Details appear in Chapter 3.

Analyses were conducted on paired data, that is, learner records were only retained if each one had a baseline and final score. This may have resulted in attrition bias which was investigated, and the balance statistics appear in the annexures. The findings confirmed that no such attrition bias occurred. Hence, the paired analyses were continued and underpinned the results summarised above.

### **Findings pertaining to buy-in, ramp-up and up-scaling**

Establishing an intervention such as this took much time and effort, in some cases more than was anticipated. Nevertheless, firm foundations were created through diligent attention to communication structures, buy-in strategies and activities, drawing up detailed work plans and resourcing the programme properly. These foundational components of the implementation will require ongoing attention and continued motivation among all participants. This will include refresher training and regular communication and information sharing, especially as new cohorts and schools are incorporated in the programme. Up-scaling commences in Year 2 and will be studied and commented on more meaningfully in future. It was observed that widespread poverty in the community and problems with school functionality at some sites had to a certain extent complicated intervention implementation and evaluation data collection.

### **Findings pertaining to monitoring of achievement testing**

The project design provided for how learner assessment tasks would be shared between the intervention and evaluation teams. Part of this arrangement involved enlarging the assessment monitoring (quality assurance) component. Approximately 30% to 40% of the assessment events at the baseline, 20% to 50% at the mid-term (end of Semester 1, only project schools), and as high as 50% to 65% during November, were monitored, depending on grade level and school sub-group, as well as practical and logistical factors related to setting up field-visit schedules. Monitoring included activities related to scoring tests and recording marks in addition to standardised assessment administration. Qualitative information from the monitoring sheets revealed that test administration standards were high. Testing followed the standardised administration manual tightly. Quantitative comparisons of learner achievement gains and baseline and end-of-year achievement levels between monitored and not-monitored assessment yielded inconsistent and even contradictory, and very often counterintuitive, results across project and control schools. "Counter-intuitive" refers to the fact that observed effects were actually contrary to what would have occurred had the service providers influenced assessment processes or marks unconsciously or systematically. Discussions in the text and figures in the annexures elaborate on this further. Achievement testing and scoring are considered to have been of high quality and free of systematic bias. Details of these analyses and comparisons appear in Section 3.3.

### **Concluding remarks and recommendations**

The evaluation methodology and data demonstrated that project-school learners benefited strongly in terms of the direct objectives of the intervention. Monitoring the learner-assessment activities also provided much support for the credibility of the assessment and evaluation data and outcomes. The relevant information and findings support continuing the interventions among the first cohort of recipients, now in Grades 2 and 5. Commencing with interventions to the second cohort of Grade 1, 4 and 7 learners is also supported. Maintaining all consultations and communications with every stakeholder to ensure that the intended benefits reach the recipients through well-informed and highly-motivated personnel remains crucial. Related to this would be ongoing attention, as had been successfully achieved in 2013, to refresher sessions with intervention staff to ensure that implementation and assessment quality remains at the current high levels. Continued tracking in the medium term of everyone's progress in the programme and undertaking the related monitoring on an annual basis remain critical for evaluating the impact of the intervention and for reaching firm conclusions about the consistency and sustainability of the observed achievement gains over a longer period of time.

## **Introduction: The origins of the language intervention**

On 17 June 2011 the United States Agency for International Development (USAID) issued a call for concept papers on the basis of its then current Annual Program Statement (APS) No. 674-11-00045. The call was issued from within its Southern African Regional Acquisition and Assistance Office. USAID's intended support would be formalised within its School Capacity and Innovation Program (SCIP). USAID was seeking to expand its bilateral work in basic education in South Africa. As a result, it invited applications from local organisations involved in promising, innovative and (partly-)tested approaches for building the capacity of school-based educators.

The non-profit or community-based organisation, siyaJabula siyaKhula (sJsK), had been working for a couple of years on such a community-supported language learning and teaching intervention in Gauteng. The first results from monitoring and evaluation by sJsK internally and by the CSIR provided reason for optimism. sJsK's interventions seemed able to regenerate learners capabilities/skills in relation to large and growing conceptual gaps that they had been suffering through participation all along in the broader schooling system.

Another HSRC study, the conceptual foundations of which sJsK had also been participating in for a short while at that stage, also paved the way for the present intervention and its evaluation. It was a consortium study focusing on literacy teaching and learning in Foundation Phase (Grade 1-3) classrooms. It set out studying classroom interaction through in-depth qualitative analysis at the time.

A further key aspect of USAID's call for concept papers was that the intended intervention should be implemented and evaluated with a view to its scaling-up to the greater system provincially and/or nationally. To satisfy the requirements set by USAID for a formal impact assessment component of the intervention, sJsK and HSRC pooled their efforts.

The sJsK-HSRC partnership subsequently went through a prescribed two-stage application process. First, an initial concept paper was submitted. Second, on favourable review, an invitation was extended by USAID to compile a full application. A key requirement was to demonstrate "promising initial results in improving learner outcomes by building teacher effectiveness and strengthening classroom and school management." USAID's approach focused on the expansion and refinement of proposals during their adjudication. This also involved support after the shortlisting process, aimed at improving the quality of proposals further.

Towards the end of the first semester of 2012 USAID, with another funding agency it had in the meantime brought on board, the Elma Foundation, contractually committed to supporting the intervention and evaluation work of sJsK-HSRC. Further support then followed during a ramping-up period straddling the second semester of 2012. The implementation proper could thus begin early in 2013. Much of the negotiations up to and just after awarding the funding grant involved conceptualising inputs and other support from all the relevant programme managers from the funding and development agencies.

## **1. Background**

### **1.1 Ethical clearance of the evaluation project**

All HSRC research has to be cleared by its Research Ethics Committee. In-principle approval of the broader design and methodology of the study was obtained early on. This was followed, before engaging with any research participants, by the development and submission of the necessary information sheets, consent forms and data-collection instruments. These comprised learner achievement assessment instruments and background information questionnaires. The latter were intended for the circuit manager, school principals, teachers and a parent or caregiver of each learner from the project and the control schools (see design

features further below). The various instruments and forms were cleared. Participants were promised anonymity and confidentiality and assured that their participation was voluntary. The principles of being treated with respect and not being exposed to risk of harm were enacted in this way.

## **1.2 Literature**

Key literature was cited in the technical proposal in support of the funding application. Such references are not cited again in this report. The initial argument focused on two central issues. The first was highlighting the relevant aspects of the current crisis in schooling, especially concerning literacy and language teaching and learning, and learner achievement. The second was the learner regeneration thrust that would underpin the envisaged intervention programme. The foregoing was accompanied by explanations about the procedures and systems that would enable successful delivery. Learner, teacher and system capacitation form the three pillars of the programme.

In essence, literature identified and consulted on an ongoing basis has supported USAID's contention that something had to be done to solve the problem of education quality failing South African learners. Poor conceptual learner foundations were considered critical and needed to be addressed. The USAID's proposal call was backed by argument and information from various national (DBE) and international (OECD) documents, such as survey outcomes, strategic plans, delivery agreements, budget-vote speeches and official statistics. These sources and evidence indicated that large investments in education and access to initial schooling have not yet led to the provision of quality education.

USAID acknowledged that factors such as inadequate infrastructure in schools, high HIV/AIDS rates among employees, underperformance among teachers and problems related to school management prevented the country from reaching the ideals of education as a national priority at a broad level. Hence, the low achievement of South African learners on internationally comparative tests comes as no surprise. Neither do their low marks in the internal systemic evaluation programme (the former Systemic Evaluations, existing Grade 12 results and the new Annual National Assessments). A central conclusion was that teaching and learning should improve. Suggested avenues include: addressing limited content knowledge, poor didactic practice and low morale among teachers, and broader systemic ailments pertaining to expectations of mediocrity, lack of teacher accountability, weak school leadership and ineffectual district/circuit support. On the bright side, pockets of excellent learner/school achievement amidst low resource levels hold much promise. It appeared that strong school management, effective classroom practice and the implementation of specific methodologies or innovations would carry the day. It was accepted that such local insights had to be supplemented by wider international knowledge about improved professional development and interventions among school educators in order to drive improved learning outcomes. In these, teacher effectiveness through sound classroom management and instructional leadership through strong school management were considered pivotal, not forgetting district monitoring and support. In response, USAID issued their SCIP call to identify and expand promising pockets of innovation in the schooling system in South Africa, in line with USAID's strategy of focusing on reading improvement among learners in primary grades.

## **1.3 The evaluation brief**

In a manner that integrates intervention implementation and impact evaluation, the proposal call determined the following:

"The overall purpose of the School Capacity & Innovation Program (SCIP) is to improve learning outcomes by building teacher effectiveness and strengthening classroom and school management. SCIP intends to invest in innovative, local interventions that will positively impact learning outcomes, as measured by improvements in primary grade reading. USAID does not have a single, prescribed

approach, but recognizes that effectiveness can be enhanced in a variety of ways including improvements in content knowledge, instructional practice, management capacity and/or morale.

Models must not be completely untested, but rather have been implemented with promising results, as evidenced by improved learner performance and/or increased demand for their services by schools and districts. SCIP seeks models or interventions that desire to further expand and have a vision for growth, including implementation at the primary school level. Applicants must be prepared to refine and rigorously test their model so that there is clear evidence of its cost effectiveness and impact in order to gain broad and sustained public support.

USAID seeks programs that can progress from implementation at a school or district level to the provincial or national scale. USAID will provide support in order to transition programs from current operating levels through a process of expansion, refinement and evaluation in anticipation of broad-based scaling. USAID intends to work with government to promote the commitment to and capacity for scaling by provinces and/or national departments. USAID is working to develop partnerships with other donors in order to develop a multi-donor network to support this initiative.

In addition to seeking initiatives that demonstrate innovation, impact and a vision for growth, USAID will give special consideration to entities that demonstrate existing partnerships with school districts and provincial departments of education. USAID will also give special consideration to applicants who draw on the respective expertise of a partner organization or organizations (e.g., NGOs, universities, for-profit ventures) in order to enhance overall program effectiveness and efficiency.”

As a result, the joint sJK-HSRC technical application / proposal listed a number of key components that would form part of the evaluation rationale, plan and objectives. They included:

- Determining which baseline achievement and contextual data to collect and how best to do so in order to demonstrate intervention impact. The sub-fields of proficiency and background information that were decided on are reflected in the reporting section and not referred to further here. Comparison against a control group is briefly explained further in the methodology and design section.
- Relating intervention focus and volume all along with the needs apparent from contextual conditions and learner achievement gains over time. Determining the appropriate indicators and evidence would be the joint task of the intervention implementation and evaluation team through regular interaction with other stakeholders according to an annual rhythm.
- Paying special attention to the attribution of impact to intervention activity. This would rest firmly on three requirements. First, properly operationalised and quantified baseline data, followed by proper mid-term and summative tracking of learner achievement and relevant related information. Then, a good overview of intervention contents and volumes. Third, equivalent scores and information are collected for sufficiently comparable control sites.
- Flexibility to allow for regular implementation monitoring of the interventions by the implementation service provider while retaining relevance, validity, and sufficiently independent or “external” evaluation. Indicative quality assurance mechanisms were identified where relevant.
- Interaction with programme beneficiaries and constituents through informal (*ad hoc*) and more formalised (regular or planned) forums or platforms.
- Alignment with international and national benchmarks and achievement targets (e.g., ANA, PIRLS, Action Plan 2014).
- Aligning the intervention project management purposes related to monitoring and tailoring intervention roll-out with external monitoring and evaluation. Nevertheless each may implement some unique but also common instruments and procedures through a shared management approach and making use of some internationally used and standardised assessments.
- Keeping an eye on resource, management and capacity challenges related to replication to scale, especially by involving and tracking cohorts over three years, adding schools every year, and expanding the layers of hierarchy related to implementation, training and supervision.

In short, the separation between and alignment of intervention implementation and sufficiently independent evaluation (M&E) components are both evident above.

To reiterate, this is the first formal evaluation report. However, it is a substantive refinement of approaches and procedures “tested”, as it were, during production of a provisional internal report generated after one semester. Evaluation procedures, data management and data analysis have been honed throughout.

As with other anticipated enhancements to the evaluation as the study progresses, the following components will also be considered more meaningfully in later reports, if not in the final report only:

- making reference to or factoring in the influence of intervention quantities / dosage,
- determining the influence of contextual factors on different rates of improvement in learner achievement between learners in project and control schools, and
- final cost-effectiveness outcomes (although some preliminary figures appear in this report).

It is anticipated that the effects from any or all of these components will become apparent once sufficiently stable and continuous intervention implementation has occurred, including a second year of expansion.

As this report is only about the impact evaluation, the focus henceforth is on the evaluation approach and methodology, and not the broader intervention programme at all. The specifications for and contents of the latter are documented well in the 13 September 2011 second amendment to the SCIP proposal call, and the subsequent full proposal (technical application) submitted by sJsK and accepted by USAID / Elma Foundation. These include matters such as the relevant contextual analysis, programme hypothesis and theory of change, (intervention) programme goals and objectives, methods, activities and implementation plans. These aspects were all aimed at the desired learner, teacher and system capacitation, as expressed in the proposal / application.

## **2. Methodology and design**

### **2.1 Evaluation purposes / objectives**

As already evident from the previous section, the HSRC has to determine the impact of the intervention. It also has to study the influence of some circuit- and school-uptake factors and systemic issues on scaling-up any successful intervention.

The key components cited in the previous section (1.3; see bulleted points) directly address the requirements for the intervention’s monitoring and evaluation plan as set in the second amendment to the SCIP proposal call. In short, they pertain to: the types of baseline data that must be collected to demonstrate impact; the process by which indicators are identified, monitored and used to enhance project management; the attribution of impact at the end of the activity; measurement and data management methods, sources and frequencies; response to the concerns of programme beneficiaries / constituents and other stakeholders; pursuit of national and international strategic goals for improving reading skills; data reliability, validity and effectiveness (in collection and analysis); and continuous modification of the intervention implementation through evaluation and assessment to ensure replication at scale.

Besides the generic monitoring and evaluation plan that formed part of the technical application / proposal submitted by sJsK, as referred to so far, a separate “HSRC Indicator Framework and Monitoring and Evaluation (M&E) plan” was compiled immediately on being awarded the grant to complement the annual work plans developed by sJsK. This more detailed document stipulated some of the further evaluation details on which information in the following few sub-sections are based.

## 2.2 Approach to the study

The study has an experimental- and control-group design. That entails a group of schools, similar in as many main respects as can be accounted for, serving as the counterfactual situation. This would allow using statistical probability methods to determine what would have prevailed had interventions not taken place. Because schools or villages were not assigned randomly to the two groups, the study is quasi-experimental. A further outcome artefact rests on involving learners from relatively small numbers of intervention and control schools. This would require taking cognisance of the effect of the feature of clustering during evaluation. Interpretation of the findings should therefore achieve a sound balance between the apparent value, meaning and usefulness of quantitative information and the resulting seemingly exact statistical indications of significant impact, on the one hand, and qualitative interpretation of proficiency gains among learners from intervention schools as opposed to learners from control schools, on the other hand. The latter would focus more on qualitative interpretation of what would be meaningful effect sizes.

The statistical approach used to determine impact is known as double-difference or difference-in-difference analysis or comparison. A simplified schematic depiction of the model is shown in Figure 1. The “location” of the difference-in-difference statistics is shown in the illustrative scheme depicted in Figure 2.

Group	Pre-test	Intervention Semester 1	Mid-term test	Intervention Semester 2	Post-test
Control (C)	$O_c$	-	$O_{1c}$		$O_{2c}$
Project (X)	$O_x$	$X_1$	$O_{1x}$	$X_2$	$O_{2x}$

Figure 1: Simplified schematic view of quasi-experimental impact evaluation design

Baseline (Pre-test)	Intervention Semester 1	Mid-term test	Difference 1 (@ base)	Test at* end Year 1	Difference 2 (@ base)	Grade 5 testing #	Difference 3 (@ base)
Gr 4 Lit.		Gr 4 Lit		Gr 4 Lit		Gr 5 Lit	
E (45)	YES, yes, y	E (58)	13	E (60)	15	E (55)	10
Same/diff baseline		diff from baseline		diff from baseline		diff from baseline	
C (42)	No	C (45)	3	C (45)	3	C (37)	-5
	Diff-in-diff scores	(in % pts)	10		12		15

\* After Semester 2 intervention.

# After another year of interventions.

Figure 2: Illustrative schematic view of the result of difference-in-difference analysis for one cohort

## 2.3 Participants / sample

A central decision in the intervention implementation and evaluation was to have a community-based study. In addition, the implementation challenges, possible solutions, scaling up and evaluation had to be appropriate to “deep”-rural conditions and the assumed limitations brought about there by longstanding disadvantage and resulting backlogs. During conversations between sJsK and the funding agencies it was decided not to base the study in relatively well-resourced, albeit underperforming, and rather well-researched schools in Gauteng. As a consequence, a complete set of villages within the demarcated part of a local tribal authority in Limpopo would receive the implementation. Indications of the deprivation and challenges applicable there are the 85% unemployment rate, and assumed high inability of parents and caregivers to support their children’s education optimally. The latter would potentially result from parents’



inability to read and write well at the levels required for helping their children. This would already apply to parents' African-language mother tongues. However, the situation would in addition be aggravated much in relation to many parents' lack of mastery of English as the eventual school language. A control group was selected from villages adjacent to the tribal-village area. On face value, similar socio-economic conditions and other contextual circumstances prevailed everywhere. A specific section (3.1) is devoted to summarising the so-called balance statistics in support or refutation of a claim for baseline similarity.

Once these village areas had been selected, a complete population of schools, teachers and learners, as it were, would participate in the intervention implementation and evaluation. Three cohort cycles are involved. The current evaluation is about the first cohort of learners in Grade 1, 4 and 7 in 2013 who reached the end of the first of three years of involvement. A single service provider, sJsk, is responsible for the complete intervention implementation. Outside the present first-year evaluation this cohort continues receiving intervention implementation as Grade 2 and Grade 5 learners in 2014 and as Grade 3 and Grade 6 learners in 2015. In 2014 a second cohort of Grade 1, 4 and 7 learners commences with a third following in 2015. The former includes the addition of more project and control schools from the existing Malamulele North-East Circuit and the additional Vhumbedzi Circuit. Interventions terminate during 2015.

The eventual sample realisation, or completion rates, are not reported here, but in the findings section (3.2) to avoid repetition. Suffice it to say that interventions during the first year involved all Grade 1, 4 and 7 learners and their teachers from 11 project and 5 control schools. The numbers of schools and villages were almost equal. More than one school per village was the case only when schools were split into junior and senior primary schools. Schools on average had about 100 learners per grade. Year groups of fewer than 40 to 50 learners were the exception, as were extremely large groups of over 200.

## **2.4 Instruments**

Two main types of instrument were used in the evaluation. They are background or contextual questionnaires and a range of learner achievement assessment tools. The former were purposively developed by the HSRC. The latter were selected by sJsk and approved by the HSRC.

### **2.4.1 Background questionnaires**

The development of the contextual questionnaires followed a two-stage process. Based on many similar previous assignments, and on this project's intervention objectives and relevant indicators, the HSRC team developed and proposed a concept instrument for each of four contexts. These comprised the circuit with jurisdiction over the schools, the schools themselves, all the teachers at the schools and the learners in these schools. These instruments were then commented on by the sJsk team members in the field. On the basis of years of exposure in similar environments, they provided excellent guidance on the reading level (register) of and length of (effort required by) the draft versions. The final instruments were thus greatly simplified, shortened and focused on only essential topics. A set of very concise and manageable questionnaires was the outcome.

Again, to avoid repetition, the content components are not reported here, but only introduced in the findings section (3.1) where equivalence between the project and control schools is reported on.

### **2.4.2 Learner assessments**

The nature of the intervention implementation and the sheer scale of the learner assessment work determined that a joint service provider and evaluator responsibility was agreed on.

The intervention needs determine that the service provider has to adjust and tailor-make intervention contents and learning materials to the foundational levels and subsequent growth trajectories of individual learners on a weekly to monthly basis. Learners, or small groups of 8 to 10 of them at most, require differential pacing and customisation of the material driving development of their next conceptual tasks. The intervention implementation therefore also relied on regular weekly and monthly feedback after intervention sessions and not only semesterly or annual testing.

In terms of evaluation independence, because testing and scoring were done in the field on an ongoing basis, semesterly quality assurance and monitoring were done on a large sub-sample of sJsKs data collection. Comparisons are also reported by grade level towards the end of the report (Sections 3.3.1-3.3.3) for the achievement levels and improvement rates between monitored and not-monitored learners.

Quality assurance or monitoring forms were completed on the basis of observations made by HSRC staff during the February baseline learner assessments, June assessments after the first semester, and end-of-year assessments during October/November 2013. Findings pertaining to these also appear towards the end of this report (Section 3.3).

## **2.5 Evaluation procedures**

Base-lining the context (circuit, school, teacher/classroom, learner/home) was undertaken by full-time HSRC research staff and post-graduate interns. During October / November 2012, on the basis of decent expectations that the Grade 3 and Grade 6 learners of the preceding year would end up in the Grade 4 and 7 cohorts of 2013, the mentioned two grades were covered mostly. The remainder were covered early in 2013 as part of the now Grade 4 and 7 groups.

The Grade 1 cohort was base-lined early in 2013. This proved to be a good decision given observations about how long it took for Grade 1 classes and learner registrations and allocations to settle, and how vastly different the numbers were when extrapolated or estimated either on the basis of the Grade R group of 2012, or 2012's Grade 1 group.

Background information was collected by means of self-report questionnaires at all levels. However, an assisted process was followed in the case of parents (including regular caregivers) for the provision of learner home context information. This mainly entailed that the parents were invited to the school, where they were assisted by school staff (teachers) to complete the instrument. In some cases those requiring assistance were helped individually or in small groups, while others who were able to progress without assistance, did so.

Baseline, end-of-first semester and end-of-year learner achievement assessments were administered by sJsK. At the same time, the HSRC conducted a higher than usual portion of monitoring or quality assurance visits. Instead of the customary 10% sample, the target was set at monitoring between 30% and 45% of learner assessments. The outcomes varied between the baseline and after each semester, by grade level and whether learners were from project or control schools. These figures are reported on towards the end of this report where learner achievement levels and gain scores are compared on the basis of being monitored or not (Section 3.3). Only selected references are made to insightful observations from analysis after the first semester, because the current report focuses on findings based on gain scores from the baseline to the end of Year 1. The HSRC also assisted sJsK to compile a standardised assessment administration manual according to which facilitators were trained by sJsK whose personnel conducted all learner assessments accordingly. The HSRC's analyses were based on this data collection.

The spread of tests and their broad contents are again, in the interest of space and to avoid repetition, not reported here, but further below in the findings section (3.2).

### 3. Evaluation findings

In this section, four main sets of findings are covered. The discussion pertains to: contextual equivalence between project and control schools (3.1); the impact of the first year of intervention implementation (3.2); comparing learner achievement marks on the basis of assessment sessions having been monitored or not (3.3); and uptake factors and systemic issues influencing roll-out and scaling up (3.4).

#### 3.1 Contextual background and its equivalence between project and control schools

Background information had been collected by February 2013 before interventions commenced. Analyses were done to determine if the contexts of project and control schools were sufficiently equivalent to rule out the effect of bias when ascribing anticipated achievement gains among learners in project schools to the project interventions. This task was already completed by the time that the provisional first-semester report was produced. In order to allow a wider audience access, an abbreviated version of the initial findings is included in this report. Broadly speaking, the findings do not support a need to pair off or match project or control schools before comparing learner outcomes. The profile of the two groups of learners, teachers and schools is adjudged to be similar enough to treat them as intact control and project groups. The main evidence and arguments for such a claim follow below.

Similar comparisons will again be made in relation to the learners, teachers and principals from the new control and project schools that were recruited to form part of Cohort 2 from 2014 onwards. Six new project schools and one new control school were added from Malamulele NE Circuit and ten new project and three new control schools from Vhumbedzi Circuit. An obvious difference comprises the latter schools' use of Tshivenda as instructional language in the lower grades, the effects of which will be kept in mind during future analyses.

The purposes of collecting sufficient background information about schools, teachers (classrooms) and learner homes are threefold. First, this structured information serves as the basis on which the intervention implementation and evaluation over the three-year course of the programme can be interpreted and customised as needed. Second, perhaps as part of the summative evaluation, or for technical or academic journal articles, such background information underpins covariate analysis of the factors associated with learner achievement levels as such, or of gain scores over time. Third, a decision about pairing or matching learners from control and project schools is substantiated on the basis of the background information.

The third and first components, namely contextual equivalence and understanding the context well, are relevant at this stage of the evaluation and will be focused on next.

Analysis of frequencies and cross-tabulations were used in the case of nominal variables, while variance analysis (ANOVA) was used with quantitative constructs or variables. In the latter case, a few selected (statistically significant) F-statistics and probability (p) values are reported when indicative.

Readers should note in advance the relative importance afforded to information of quantitative and qualitative nature in this report. The latter is favoured. The reason for this is that learners cluster into 11 project and 5 control schools. Learners all come from the same geographical area and live under similar socio-economic conditions. Such a sample is not considered large, even though there may be 1 000 or more learners per grade-level. It was also not meant to represent the country beyond reflecting the situation and challenges faced by our most marginalised learners. Statistical significance testing may easily give a sense of over-confidence in reported findings. Hence, although such statistics are reported in places, they should be treated circumspectly. The most meaningful interpretations would relate to how constructs have been operationalised, and what different score levels or "measures" say qualitatively about any relevant effects. Also, statistically correcting for the effect of clustering of individual respondents into teacher and school

hierarchies at a higher level would increase standard errors (and confidence intervals) of scores as much as fourfold. This, in turn, would make it difficult or impossible to detect sub-group differences statistically, although they may be highly informative. It was therefore decided to reserve such more technical analyses for publication through articles in scholarly journals.

### 3.1.1 School equivalence

The school-level background questionnaire, completed by school principals or their designated substitutes, focused on: (a) staff capacity and responsibility (especially related to language teaching), and (b) issues relating to infrastructure, facilities and conditions at schools.

The rural nature of the area – schools are all very far from the largest regional town, Thohoyandou – and the high levels of poverty (85% unemployment) in the region, predict a consistent so-called floor effect. This means that, irrespective of being a project or control school, amenities and conditions would be quite basic. Neither learners from project nor from control schools as a group would in a systematic way have greater access to favourable learning dynamics at school level. The evidence below broadly supports not pairing or matching specific project and control schools on the basis of different levels of conduciveness in school conditions for language learning and teaching before evaluating intervention impact.

Eleven project schools were identified in advance. After the April 2013 holidays, a 12<sup>th</sup> school was opened to absorb an overflow of children from three existing schools that were bursting at their seams. They, therefore, had already commenced with the programme. One of the original eleven schools, though, was reassigned as a control school by sJSK because its readiness for the intervention programme suffered extended delays. That brought the number of control schools to five, instead of the intended four, but reduced the project schools again to 11.

#### *Staff capacity and responsibility*

A summary is provided in Table 1 of the staff capacity situation as experienced by project and control schools. Only the number of vacancies for non-language teachers was noteworthy (significant at the 10% probability level). Other than that, rather similar conditions prevailed at project and control schools. Conditions in project schools were slightly better than in control schools.

**Table 1: Capacity components favouring project and control schools respectively**

<b>Conditions rated more highly in project schools</b>	<b>Conditions rated more highly in control schools</b>
Schools have full staff complement ( <b>55%</b> > 20%)	Number of principals (0,9 < <b>1,0</b> )
Number of SMT vacancies per school ( <b>0,4</b> < 0,8)	Number of language teachers in IP (3,4 < <b>4,2</b> )
Language teacher vacancies per school ( <b>0,5</b> < 1,0)	Number of language teachers - Gr 7 (1,0 < <b>1,6</b> )
Number of other-teacher vacancies ( <b>0,4</b> < 1,4) <sup>@</sup>	Number of other teachers - Gr 7 (2,4 < <b>4,0</b> )
Number of Deputy-Principals ( <b>0,7</b> > 0,4)	Nr of administrative/support staff (0,2 < <b>0,4</b> )
Number of Language HoDs ( <b>1,6</b> > 1,4)	
Number of HoDs in other subjects ( <b>1,6</b> > 1,2)	SMT = School Management Team
Number of language teachers in FP ( <b>6,4</b> > 5,4)	HoD = Head of Department
Number of other teachers in FP ( <b>3,1</b> > 2,6)	FP = Foundation Phase
Number of other teachers in IP ( <b>5,6</b> > 4,6)	IP = Intermediate Phase
<b>No substantive difference or relevance</b>	
Sex of principal (male) – (73% ~ 60%)	

<sup>@</sup> p<0,10 (eta-coefficient)

The numbers of learners and classes are reflected in Table 2. No differences between project and control schools were close to statistical significance, despite control schools having slightly more senior and slightly fewer junior learners, and being slightly larger overall.

**Table 2: Number of learners and classes on average by grade for project and control schools**

	Type	Gr R	Gr 1	Gr 2	Gr 3	Gr 4	Gr 5	Gr 6	Gr 7	Total
Learners	Project	85	110	93	81	85	77	70	68	669
	Control	79	98	83	92	88	83	75	76	674
Classes	Project	1,4	1,8	1,9	1,7	1,9	1,5	1,5	1,4	13,0
	Control	1,4	1,8	1,6	1,8	1,8	1,8	1,6	1,6	13,4

### *Infrastructure, facilities and conditions*

Table 3 reflects the sufficiency and quality of a range of facilities and conditions as rated by schools on a five-point Likert scale<sup>3</sup>. Fencing / security at control schools was rated as good to very good, while the rating for project schools was average, and differed significantly between project and control schools. The only other significant difference, also only within the 5% level of probability, was the higher quintile ranking at project schools.

Conditions in control and project schools were considered as average to good for electricity, tapped water, permanent buildings, usable classrooms, flushed toilets for staff (albeit poor in project schools), photocopier/s, and computer facilities in the office and for staff. Facilities were rated as poor or very poor, if in existence at all, in the cases of an office block, flushed toilets for learners, telephones, a school hall, a staff room, a library, computer facilities for learners, and a walk-in safe room or other secure storage. Again a rather balanced distribution of conditions was evident, with slightly more of them favouring control schools this time. Given that the reported differences were not significant at the 1% probability level, and not directly language related, the two groups of schools can be considered very similar. Achievement increases among learners in project schools would then most likely result from the intervention implementation, with no further need to control for school effects by means of matching or pairing specific project or control schools.

**Table 3: Existence, sufficiency and quality of school facilities for project and control schools respectively**

Conditions rated more highly in project schools	Conditions rated more highly in control schools
Electricity ( <b>2,8</b> >2.4)	Usable classrooms (2,2< <b>2,4</b> )
Tapped water ( <b>2,1</b> >1.8)	Flushed toilets for staff (0,8< <b>1,6</b> )
Permanent buildings ( <b>2,4</b> >1.6)	Flushed toilets for learners (0,6< <b>1,0</b> )
Office block for principal and admin staff ( <b>1,4</b> >0.6)	Telephones (0,6< <b>1,4</b> )
Photocopier ( <b>2,8</b> <2.6)	Fence / security (2,2< <b>3.2</b> )*
Staff room ( <b>0,7</b> <0.6)	Hall (0,9< <b>1.4</b> )
Storage / walk-in safe room ( <b>1,0</b> >0.6)	Library (0,3< <b>0.6</b> )
Quintile ranking of school ( <b>1,9</b> >1.4)*	Computer facilities for staff / office (1,8< <b>2,0</b> )
	Computer centre / facilities for learners (0,2< <b>0,4</b> )
	Sports grounds (0,6< <b>1.4</b> )
No difference	
Language laboratory (0,0 ~ 0,0)	
No-fee schools; no children paying any fees (All schools ~ All schools)	
Have school feeding scheme with five meals a week (All schools ~ All schools)	

\* p<0,05

### 3.1.2 Teacher and classroom equivalence

Mirroring the school-level background questionnaire, the teacher contextual questionnaire was used to collect information from teachers about: (a) teacher capacity and responsibilities (especially related to language teaching), and (b) classroom infrastructure, facilities and materials.

<sup>3</sup> 0=None/very poor; 1=Poor; 2=Average; 3=Good; 4=Very good

Although broader school conditions may influence teaching, teachers bring a range of unique skills and motivations into the classroom. The extent to which teachers and classrooms differ between the project and control schools is documented briefly next.

There were 176 teacher records in the dataset. Eighty-four (84) teachers chose not to complete a contextual questionnaire. They were distributed evenly across project and control schools at an average of about five per school. Voluntary participation is a cornerstone of respect to research participants. Of the 176 teachers, 114 teachers were from project and 62 from control schools. The distribution in relation to the sex of teachers from project schools shows that 36 (32%) were male and 76 (68%) were female. The corresponding figures for teachers from control schools were 23 (38%) male and 38 (62%) female. Three teachers did not indicate their sex on the questionnaire.

#### *Teacher capacity and responsibility*

Qualification levels and experience between project- and control-school teachers are summarised in Table 4. Largely similar distributions are evident. Sometimes teachers from project and or control schools had more experience depending on phase or grade level, teaching site (present post, school, etc.) and subject taught. Teachers on average had a highest qualification level of marginally below a three-year degree. The distribution differed very little and not significantly between project- and control-school teachers.

**Table 4: Capacity comparison for teachers from project and control schools**

Conditions rated more highly in project schools	Conditions rated more highly in control schools
Overall teaching years ( <b>18,0</b> >17,8)	Overall years teaching language at IP (11,3< <b>12,1</b> )
Overall years teaching language at FP ( <b>16,9</b> >13,8)	Overall years teaching Gr 7 language (9,6< <b>9,7</b> )
Years teaching IP in present post ( <b>10,1</b> >8,9)	Teaching FP language in present school (14,6< <b>16,8</b> )
Years teaching Gr 7 in present post ( <b>10,1</b> >7,8)	Teaching IP language in present school (10,4< <b>10,6</b> )
	Teaching Gr 7 language in present school (7,6< <b>9,7</b> )
	Years teaching FP in present post (17,0< <b>19,1</b> )
<b>No substantive difference</b>	
Highest qualification level <sup>4</sup> of teacher – (1,86 > 1,83)	

FP = Foundation Phase  
IP = Intermediate Phase

Brief comments now follow in relation to qualifications, training, teaching loads and language proficiency among teachers. Though distributions did not differ substantively between teachers from project and control schools attention is drawn to the following aspects. Almost half of the teachers completed at least a diploma; 48% of project school teachers and 41% of control school teachers. More teachers at project schools had four-year degrees (32%, compared to 17% at control schools), while more control school teachers had three-year degrees (36%, compared to 20% at project schools) and post-graduate qualifications (5%, compared to 0% at project schools). Differences were inconsistent and insignificant across qualification levels. In both sub-groups, the major fields of study that teachers pursued in their highest qualifications were English, followed by Xitsonga and then Education Management. A majority of teachers reported having studied Xitsonga as a major in their next-highest qualification.

Teachers received the regular wide range of language and literacy teaching training as part of periodic NCS and CAPS courses, especially tied to policy changes nationally. The self-rated relevance and quality of these courses did not differ statistically significantly between teachers from project and control schools. Teachers from project schools were always marginally more positive about both the relevance and quality of such training, which were mostly rated by all teachers halfway up from “in between” to “high” (the 3-point and 4-point positions on the 5-point rating scale provided).

Teaching loads and patterns were very similar between teachers from project and control schools. Most teachers taught three subjects. A minority (about 40%) had to teach a fourth subject. Their core subjects often involved Language / Literacy (Xitsonga and English), Mathematics / Numeracy and Life Skills. These

<sup>4</sup> 0=Gr 12; 1=Diploma; 2=Degree; 3=Honours / four-year degree; 4=Master’s or doctoral degree

subjects form a package for FP anyway. Second, third and fourth subjects were frequently taught at lower grades. The number of lessons per subject per week also declined as teachers moved from their first to their fourth subjects, ranging from 14 to 16 at first, through 10 to 14, and 6 to 8, down to 4 to 6. Similarly, the time spent per week on each subject also declined from around 450 minutes through 400, and 240-360, to 180 or slightly above, especially at project schools. The number of learners per subject or class was very stable at around 40 to 48, again with more project-school teachers at the higher limits.

Of the control-schools and project-school teachers respectively, 90% and 92% had Xitsonga as home language, and 9% and 7% Tshivenda; again a nearly identical distribution. No statistically significant difference was evident in the proficiency of project- and control school learners in speaking, reading and writing their home language. Such proficiency was always rated as excellent. Other languages that teachers also spoke included English (indicated by 84% of teachers in control schools and 75% of those in project schools), followed by Afrikaans (35% and 44% respectively), Tshivenda and isiZulu (around 20% to 35% for both languages and across school types). Speaking, reading and writing proficiencies in their second to fourth languages were generally rated as good to halfway between good and average. Writing proficiency was always rated lowest, with teachers from control schools showing statistically significant higher proficiency in relation to writing English and Afrikaans as their most prevalent second and third languages (F=4,923; p=0,028; and F=4,316; p=0,040 respectively).

#### *Classroom infrastructure, facilities and materials*

The existence and quality of classroom infrastructure and facilities, rated on a five-point Likert scale<sup>5</sup>, are summarised in Table 5. Teachers from project and control schools reported working under largely similar conditions in their classrooms. Electricity and lighting were generally rated as good. Air quality, comfortable temperatures, permanent building materials and structures, desks and chairs for teachers and learners, overall space and lockable or other storage were all rated as around average, while shelving was considered to be in shortage and rated as poor. Five items showed statistically significant differences across a range of probabilities in relation to the quality ratings between the two teacher groups. These differences mostly favoured control-school teachers. However, the earlier note about the effect of clustering within a sample should be heeded. Adjustment for that will render most of, if not all, the differences insignificant, a matter that is not explored further but left for future publications of a technical nature.

**Table 5: Existence, sufficiency and quality of classroom infrastructure and facilities for project and control schools**

Conditions rated more highly in project schools	Conditions rated more highly in control schools
Enough fresh air (3,6>3.2) <sup>@</sup>	Electricity (3,9<4,2) <sup>@</sup>
Learner desks (3,0>2.9)	Appropriate light / enough windows (4,1<4,2)
Overall space (3,3>3.1)	Permanent building materials (3,3<3,8)*
Shelves (2,4>2,3)	Learner chairs (2,7<2,8)
	Teacher desk (2,9<3,5)*
	Teacher chair (3,0<3,8)***
	Lockable and other storage (2,8<3,3)*
<b>No difference</b>	
Comfortable temperature (3,0 ~ 3,0)	
* p<0,05    **p<0,01    ***p<0,001	(No corrections calculated for this report for effect of cluster sampling)

The availability of specific materials in classrooms, and their condition, rated on a four-point Likert scale<sup>6</sup>, are summarised in Table 6. Although provision and quality were always rated higher at control schools, many such differences were rather small and not significant. No differences at all were evident in a few cases. Even significant differences are of little practical meaning, and would fall away on correction for clustering in the sample design.

<sup>5</sup> 0=None/very poor; 1=Poor; 2=Average; 3=Good; 4=Very good

<sup>6</sup> 1=Poor; 2=Average; 3=Good; 4=Excellent

**Table 6: Availability and quality of materials used in classrooms for project and control schools**

Availability of materials (Conditions rated more highly in control schools)		Quality of materials (Conditions rated more highly in control schools)	
Prestik / adhesives (0,7<0,9)*		Chalkboards (2,8<3,1)*	
Wall charts (0,7<0,8)		Chalk (3,1<3,3)	
Drawing pins (0,4<0,7)***		Prestik / adhesives (2,9<3,0)	
		Wall charts (2,6<3,0) <sup>@</sup>	
		Board erasers (2,9<3,2) <sup>@</sup>	
		Drawing pins (2,4<2,9) <sup>@</sup>	
		Rulers or T-pieces (2,6<2,9)*	
<b>No difference</b>			
Availability of chalkboards, chalk, board erasers (All schools ~ All schools)			
Availability of rulers or T-pieces (0,7 ~ 0,7)			
* p<0,05	**p<0,01	***p<0,001	(No corrections calculated for this report for effect of cluster sampling)

Chalkboards, chalk, adhesives/Prestik, wall charts and board erasers were freely available, while board rulers / T-pieces and drawing pins were not. The latter two items were also rated as below average in quality, while the remainder were rated as average or slightly above.

The relatively small number of significant differences revealed above, the marginal association between many of these components and direct language-teaching issues, the inconsistent patterns of advantage in favour of either project-school teachers or control-school teachers, and the confidence or probability levels mostly being below 99% and often below 95%, together with the fact that this will further diminish when factoring in the effect of clustering, provide strong support for claiming equivalence between the two groups of schools and teachers, so as not to influence the implementation or outcomes of the intervention differently. It would be likely that achievement increases among learners in project schools beyond the level of increases obtained by learners in control schools would be the result of the language intervention.

### 3.1.3 Home and learner contextual background equivalence

The home-level background questionnaire, completed by the parents or caregivers of learners, focused on: (a) parental characteristics and language proficiency, (b) learner access to school, and (c) learner socio-economic conditions and exposure to language-related materials and behaviours. The observed high degree of school and classroom equivalence was also expected to prevail at home and in relation to learner-related characteristics and language support. Evidence in this regard is summarised next. This is done mostly in bullet (list) form under five appropriate sub-headings because the many response formats in the questionnaire could not be reflected in a consistent manner in a standard table. Also, the practical meaning and qualitative extent of differences between score distributions for project- and control-school learners' home conditions were emphasised rather than statistical comparison. A strong motivation for this choice again was that clustering of learners into 11 project and 5 control schools would tend to foster false confidence in seemingly exact indications of statistically significant sub-group differences at learner level. Originally comparative statistics were calculated, but are not reported owing to their detail and volume. Statistical adjustments were not done being considered too technical for the present report, besides distracting from an emphasis on intervention impact.

The learner context baseline dataset comprised 2 904 records. A few parents (n=65) did not complete the field for the sex of learners. Of the remaining 2 839 learners, 52% and 49% respectively for project and control schools were boys. Thus, the distribution was very similar across type of school.



### *Parental characteristics*

- Duration of family living in their village: 29,4 years for the project group; 22,0 years for the control group. Comment: Families attached through heritage to the Mhinga tribal villages, from which the project schools were selected, may experience greater stability and less mobility than those from control schools, selected from the area just west of Mhinga. However, families from both areas had often been living there for more than 20 years. It is assumed that learners from both areas therefore experienced community and home life similarly, which is unlikely to affect their learning at school.
- Who learners live with (during the week, every day): At least 70% of learners lived with either or both parents, with the mother being indicated by a large majority; another almost 10% of children stayed with their grandmothers. Comment: Slightly higher percentages (4-9 %-points) of learners living with either or both parents were recorded for control-school learners, compared to project-school learners. This should be inconsequential, also in view of relatively uneven questionnaire completion and data quality.
- Age of parent/s or caregiver/s: A third were 30 years of age or below (including almost 6% below 20 years in age), another third up to 40 years of age, around 20% up to 50 years of age, 10% up to 60 years of age, and about 5% over the age of 60. Comment: The age distribution among the two sets of parents (caregivers) was very similar, never varying by more than 3 %-points.
- Highest qualification of parents or caregivers: No schooling for almost 13%, primary school for around 30%, Grade 8 to 9 and Grade 10 to 11 for another almost 20% each, matric for around 15%, and any post-school qualification for 2% to 3% ( $p < 0,01$ ). Slightly greater proportions of parents / caregivers of children from control schools (1-4 %-points) were better qualified from and beyond Grade 10.
- Earning capacity of parents / caregivers: No income for 17%; social grants for 70%; salaries for 10%; an income from farming and from business or trade for 1% and 2% respectively. Slightly more of those caring for children from control schools earned their own salaries (4 %-points), while more of those caring for children from project schools received social grants (5 %-points). Comments: The magnitude of these differences is too small to have practical consequences in relation to learning at school. Participants could select more than one option and "Other (specify)", but rarely did.

### *Language profiles and proficiency*

- Language spoken by family at home: Xitsonga for 99%. Comment: No sub-group differences.
- Language spoken by learner at home: Xitsonga for 99%. Comment: No sub-group differences.
- Learner proficiency<sup>7</sup> in the home language:
  - Speaking: 3,4 in project group; 3,6 in control group. Comment: Xitsonga is spoken very well.
  - Reading: 2,8 in project group; 3,0 in control group. Comment: Xitsonga is read well.
  - Writing: 2,9 in project group; 3,0 in control group. Comment: Xitsonga is written well.
- Parents'/caregivers' proficiency in the learners' home language:
  - Speaking: 3,5 in project group; 3,6 in control group. Comment: Xitsonga is spoken very well.
  - Reading: 3,1 in project group; 3,2 in control group. Comment: Xitsonga is read well.
  - Writing: 3,1 in project group; 3,2 in control group. Comment: Xitsonga is written well.
- How often<sup>8</sup> learners speak English at home: 1,9 in project group; 2,2 in control group.
- Learner proficiency in English (Likert rating scales are as above from here onwards):
  - Speaking: 1,8 in project group; 2,1 in control group.
  - Reading: 2,0 in project group; 2,2 in control group.
  - Writing: 1,9 in project group; 2,3 in control group.
- How often parents / caregivers speak English at home: 1,9 in project group; 2,2 in control group.
- Parent / caregiver proficiency in English:
  - Speaking: 2,0 in project group; 2,4 in control group.
  - Reading: 2,2 in project group; 2,5 in control group.
  - Writing: 2,1 in project group; 2,6 in control group.

<sup>7</sup> 1=Not at all well; 2=Only a little; 3=Well enough; 4=Very well

<sup>8</sup> 1=Never; 2=Rarely; 3=Sometimes; 4=A lot / often

Overall comment: The differences above are mostly marginal and assumed to be of little practical consequence. Siswati, isiXhosa and, less so, Tshivenda were reported as the home languages of families in isolated cases only. The average ratings of learner and carer proficiencies were slightly higher for control-school learners than project-school learners. Ratings of carers' reading and writing proficiencies were slightly higher than those of the learners. Learners and parents / caregivers were reported to rarely speak English at home, and generally to develop no more than a little (or basic) proficiency in this language.

#### *Learner access to school*

- How far<sup>9</sup> learners have to travel to school: 1,7 (scale points) in project group; 1,8 in control group. Over 80% of learners stay within three kilometres from school. The average would be between two and three kilometres. Fewer than 5% have to travel more than six kilometres. Comment: Slightly more project-school learners stay closer to school (3 %-points).
- How learners get to school: By far most learners walk to school (95%). Those not walking get taken to school by car (not more than 5%), or go by taxi (not more than 2%). Very few make use of bus services (< 0,5%). Comment: Slightly more project-school learners walk to school (5 %-points), while more control-school learners get taken by car (3 %-points).
- How long it takes learners to get to school: A third of learners get to school within 10 minutes, while an additional 55% get there within half an hour. Fewer than 4% require more than an hour to get to school.

Access to school as factor is not expected to significantly or practically make a difference in the project in relation to how much project-school learners would benefit from the language intervention.

#### *Learner socio-economic conditions*

A summary is provided in Table 7 of how often learner homes were reported to have the listed items.

**Table 7: Percentage of parents / caregivers indicating that learner homes have the listed items**

Item	Group		Item	Group		Item	Group	
	Project	Control		Project	Control		Project	Control
**Tap water	56,5	49,8	Radio	73,9	74,0	Satellite dish	18,0	16,3
***Electricity	90,0	83,3	@Computer	11,2	9,0	Television	82,4	80,1
***Flush toilet	17,3	6,4	*Laptop	5,4	7,8	*CD/DVD player	75,1	71,5
**Fridge	80,0	75,5	*Internet	5,7	8,3	Car	19,6	21,5
Telephone	14,9	14,6	@Cellphone	90,8	92,7	Bicycle	31,1	32,3

\* p<0,05    \*\*p<0,01    \*\*\*p<0,001    (No corrections calculated for this report for effect of cluster sampling)

@ p<0,10

Facilities and commodities in ample supply at homes include electricity, fridges, radios, cellular phones, television and CD/DVD players. From these items, with the exception of cellular phones, the homes of learners from project schools were slightly better off. About half of the homes had tapped water; more so in the case of project-school learners' homes. The remaining commodities were all in low supply, often in favour of control-learner homes. The inconsistent distribution may rule out any advantage for either subgroup.

#### *Exposure to language-related materials and behaviours*

- Learner has access to a public library: 13,5% in project group; 20,5% in control group.
- Learner has access to a dictionary at home: 27,7% in project group; 33,4% in control group.
- How often<sup>10</sup> learner reads the following materials at home:

<sup>9</sup> 1=Less than 1 km; 2=1-3 km; 3=3-6 km; 4=More than 6 km

<sup>10</sup> 1=Never / almost never; 2=1-2 times a week; 3=3-5 times a week; 4=Almost daily / daily

- Books: 2,2 in project group; 2,4 in control group. Comment: A third of learners in each group never or almost never, and once or twice a week read books; this applied more to project-school learners (with 6 and 1 %-point/s more respectively). Only 2% of learners read daily.
- Magazines: 1,3 in project group; 1,6 in control group. Comment: Almost three-quarters of project-school learners and 61% of control-school learners were said never or almost never to read magazines. Only about 4% read daily.
- Newspapers: 1,3 in project group; 1,5 in control group. Comment: Just more than 80% of project-school learners and 63% of control-school learners were reported never or almost never to read newspapers. Only about 4% read daily.
- How often<sup>11</sup> learner is given help by at least one parent / caregiver in relation to the following:
  - Homework: 2,8 in project group; 3,0 in control group. Comment: Almost half the control-school learners and 41% of project-school learners reportedly were helped on a weekly basis or more often. Forty percent received only incidental help once a month or less.
  - Information collection for home- or schoolwork: 2,6 in project group; 2,9 in control group. Comment: 40% of control-school learners and 34% of project-school learners were helped on a weekly basis or more often. Almost half obtained information once a month or less.
  - Homework completion checked: 2,8 in project group; 3,1 in control group. Comment: 54% for control-school learners and 46% for project-school learners on a weekly basis or more often. Completion of almost 40% was checked once a month or less.
  - Preparation for tests / exams: 2,6 in project group; 3,0 in control group. Comment: Almost 50% for control-school learners and almost 40% for project-school learners on a weekly basis or more often. Help was rendered once a month or less with test / exam preparation in close to half the cases.
  - By paying someone else to give such help: 1,5 in project group; 1,8 in control group. Comment: 17% of control-school learners and 11% of project-school learners were afforded this help on a weekly basis or more often. Such help was never or almost never secured in 60%-75% of cases.
  - Ensuring learner gets to and from school: 3,0 in project group; 3,1 in control group. Comment: 59% for control-school learners and 61% for project-school learners on a weekly basis or more often. This was ensured in a third of cases once a month or less.
  - Looking at learner results to improve them: 2,9 in project group; 3,3 in control group. Comment: 64% for control-school learners and 47% for project-school learners on a weekly basis or more often. Results were looked at once a month or less in 25%-40% of cases.
  - Encouragement to read and write: 3,1 in project group; 3,4 in control group. Comment: 67% for control-school learners and 56% for project-school learners on a weekly basis or more often. Encouragement was given once a month or less in 22%-32% of cases.
  - Listening to learner reading: 2,9 in project group; 3,3 in control group. Comment: 60% for control-school learners and 49% for project-school learners on a weekly basis or more often. Learner reading was listened to once a month or less in 26%-37% of cases.
  - Asking about what learners had read: 2,8 in project group; 3,1 in control group. Comment: 57% for control-school learners and 45% for project-school learners on a weekly basis or more often. Learner reading was asked about once a month or less in 30%-40% of cases.
  - Asking about learner schoolwork: 3,0 in project group; 3,2 in control group. Comment: 61% for control-school learners and 52% for project-school learners on a weekly basis or more often. Schoolwork was asked about once a month or less in 27%-37% of cases.
- For how long<sup>12</sup> learner is assisted by parent giving daily homework assistance: 2,8 in project group; 2,9 in control group. Comment: 23% of learners received as much help as needed; 10% 1-2 hours; 15% half an hour to an hour; almost 50% less than an hour, including 15% receiving less than 10 minutes.
- How many<sup>13</sup> books of his/her own a learner has at home: 1,8 in both the project and control group. Comment: Almost half were reported to have no such books; another almost 40% had fewer than five; about 10% had five to ten; fewer than 8% had more than 10 books, including 3% having more than 25.

<sup>11</sup> 1=Never / almost never; 2=Once every 3-6 weeks; 3=Once every 1-3 weeks; 4=Every 1-4 days

<sup>12</sup> 1=Very little (<10 minutes); 2=10-30 minutes; 3=31-60 minutes; 4=1-2 hours; 5=As much s/he needs

- How many<sup>14</sup> other books there are at home: 1,7 in both the project and control group. Comment: Almost half were reported to have no such books; another almost 40% had fewer than ten; about 10% had 10-25; fewer than 4% had more than 26 books, including 1% having more than 100.
- How much time<sup>15</sup> a learner spends or is required to spend at home on a typical weekday on:
  - Work in the home language subject: 2,4 in project group; 2,6 in control group. Comment: 50%-60% spent 20 minutes or less; 24%-30% spent more than 40 minutes.
  - English work: 2,1 in project group; 2,4 in control group. Comment: 58%-65% spent 20 minutes or less; 15%-21% spent more than 40 minutes.
  - Work on other school subjects: 2,2 in project group; 2,5 in control group. Comment: 51%-60% spent 20 minutes or less; 16%-24% spent more than 40 minutes.
  - Watching television: 2,3 in project group; 2,9 in control group. Comment: 35%-58% spent 20 minutes or less; 25%-38% spent more than 40 minutes.
  - Reading for pleasure: 2,1 in both project and control group. Comment: two-thirds spent 20 minutes or less; 16% spent more than 40 minutes.
  - Playing with / visiting friends: 2,8 in project group; 3,1 in control group. Comment: 25%-40% spent 20 minutes or less; 40%-50% spent more than 40 minutes.
  - Buying food/groceries for the family: 1,8 in project group; 2,1 in control group. Comment: 61%-78% spent 20 minutes or less; 12%-18% spent more than 40 minutes.
  - Doing home tasks (e.g., washing, gardening): 1,9 in project group; 2,1 in control group. Comment: 62%-73% spent 20 minutes or less; 12%-15% spent more than 40 minutes.
  - Doing own paid part-time work: 1,8 in project group; 1,9 in control group. Comment: 72%-75% spent 20 minutes or less; 15%-17% spent more than 40 minutes.

General comment: Control-group learners were consistently reported to receive slightly higher levels of exposure or support than project-school learners in relation to the relevant items listed above. The two groups to a large extent experienced very similar circumstances at home, as also at their schools. As a result, one can with confidence accept that the intervention implementation would have been responsible for any outcome differences between learners from project and control schools now observed after a year.

### 3.2 Impact of interventions

Essentially, as described further below in a more technical sense, the reporting in this section involves comparison of the pace of acceleration of learning among learners benefiting from the interventions with the counterfactual. The latter refers to what would have happened had the interventions not been implemented in the project schools. In brief, it assumes that learners in all schools should learn over time. How DBE structures schooling, prepares teachers, manages the curriculum, provides learning and teaching support materials, implements school support and monitoring, etc. take place not only in the project schools, but also the control schools. In project schools only, all the foregoing and the additional sJsK intervention take place. As a result, the anticipated different learner achievement gain trajectories can ultimately be ascribed to the sJsK intervention. Two main comparisons are made in relation to learners' achievement gains across project and control schools. Direct comparison reports on mean learner assessment score changes from the baseline scores of February 2013 to the end-of-year scores of November 2013 in the various tests. Indirect comparison investigates the growing proportion of learners who achieved at least 50% in each test across learners from project and control schools.

As indicated before, sJsK produces data and records for their own and for their funders' purposes more frequently than HSRC's semesterly evaluation reports appear. sJsK's records and analyses inform adjustments to some of their intervention components. Although slightly different approaches may be

<sup>13</sup> 1=None; 2=Fewer than 5; 3=5-10; 4=11-25; 5=More than 25

<sup>14</sup> 1=None; 2=Fewer than 10; 3=10-25; 4=26-100; 5=More than 100

<sup>15</sup> 1=0-10 minutes; 2=11-20 minutes; 3=21-40 minutes; 4=More

followed, it in no way invalidates or contradicts sJsk’s approach or figures. HSRC places a high premium on demonstrating how many sources of variance were controlled as minimum standard applicable to an independent evaluation. One mechanism employed was to conduct difference analysis only on paired pre- and post-test learner scores. This means that comparisons between baseline and end-of-year scores are based only on those learners for whom both scores were available. However, this may result in another form of bias on the basis of differential attrition of cases (i.e. those for whom only one of the two scores were available) from the project and control groups. For this reason, attrition rates and the effect of attrition are reported briefly before comparing learner achievement gains between learners across the project and control groups. Comments are also then provided on how balanced, and thus random, such attrition was. Greater consistency occurred at Grade 4 and 7 levels between the baseline and end-of-year assessments. This may have resulted from less learner movement between schools and from learners settling down quicker early in the school year at the higher grades compared to the situation in Grade 1. At Grade 1 level, many first-time decisions about school readiness and age, the best school to select and moving closer to a school, even as a result of population influx from the neighbouring Mozambique, still needed to be made. The foregoing situation explains the slight differences in frequencies or mean scores that may exist between the respective reports compiled by sJsk and HSRC.

Before reporting on learner achievement gains, some space is devoted to an overview of the intervention implementation. For possible time-series differentiation at a future stage, intervention information for the first-semester period up to the winter school holidays at the end of June 2013 and for the second semester, from July to October/November, has been kept separate thus far. This information was compiled by sJsk and provided on request.

During the first semester of 2013, the learners were exposed to six one-hour interventions at the rate of one intervention per week. These intervention sessions replaced one hour of normal First Additional Language instruction. Table 8 provides the hours of tuition applicable to learners from both the control and project schools.

The Grade 1 intervention followed a standard format for each one-hour session. This format is given in Table 9. “Instructors” refers to the sJsk-trained community members.

Because of the very low levels of literacy encountered in the area, interventions designed for learners in both Grade 4 and Grade 7 followed the same format, as can be seen from Table 10. After that, some more information is provided on the selection and training of the instructors, the intervention contents, and early indications of the cost of the intervention per learner.

**Table 8: Hours of normal and intervention tuition received by school type and grade during the first semester**

<b>Grade</b>	<b>Control schools: Normal tuition Hours</b>	<b>Project schools: Normal tuition Hours</b>	<b>Project Intervention Hours</b>	<b>Intervention %</b>
<b>1</b>	24-36	18-30	6	20%
<b>4</b>	60	54	6	10%
<b>7</b>	48	42	6	12.5%

**Table 9: Standard format of one-hour intervention sessions to Grade 1 learners in the first semester**

Item	Time	Content	Delivered by	Materials
1	10 minutes	Song in English	Instructors	None
2	10 minutes	Phonetic alphabet	Instructors	Wall and individual charts
3	10 minutes	Flash card drills	Instructors	Flash cards
4	30 minutes	Worksheet exercises	Instructors	sJsK worksheets in workbooks
<b>Total:</b>	<b>60 minutes</b>			

**Table 10: Standard format of one-hour intervention sessions to Grade 4 and Grade 7 learners in the first semester**

Item	Time	Content	Delivered by	Materials
1	5 minutes	Phonetic alphabet	Instructors	Wall and individual charts
2	15 minutes	Basic phonics drills	Instructors	sJsK worksheets in workbooks
3	15 minutes	Reading in Xitsonga and English	Instructors	sJsK bilingual reading booklets
4	25 minutes	Worksheet exercises based on memory training, comprehension, literacy and mental disciplines	Instructors	sJsK worksheets in workbooks
<b>Total:</b>	<b>60 minutes</b>			

Due to the implementation challenges faced by sJsK during the first semester of 2013, sJsK's teacher training at project schools was delayed and effectively not implemented at all yet.

Initial instructor training for community members took five weeks at three days per week, plus four weeks of in-classroom training. It took two weeks of training to achieve basic proficiency in administering and marking assessments. An additional three weeks were required for basic proficiency in teaching skills and content knowledge. In addition to this, four weeks of in-classroom on-the-job training was required to reach the required competency levels.

The intervention roll-out during the second semester of 2013 is reported next. Again, this information was compiled by sJsK and provided on request. As stated by sJsK in an end-of-year report to the funders, direct learner interventions consisted of replacing one hour of normal First Additional Language (FAL) tuition time with one hour of intervention per week (Grades 1 and 4, plus Grade 7 initially) and replacing two hours of normal FAL tuition time per week at Grade 7 level during the third and fourth terms. As a result, the Grade 1 and Grade 4 learners in the project schools received twelve one-hour interventions at the rate of one intervention per week during the second semester of 2013. The Grade 7 learners in the project schools received twenty four one-hour interventions at the rate of two interventions per week. Each of these intervention sessions replaced one hour of normal First Additional Language instruction, making a total replacement of two hours of First Additional Language instruction per week. The decision to double the intervention volume to the Grade 7 learners was taken at the beginning of the second semester to try to bridge the substantial gap between where the learners were and where they needed to be to reach

international standards by the end of the year. Table 11 provides the hours of tuition applicable to learners from both the control and project schools.

Although the Grade 1 intervention followed a standard format for each one-hour session, the difficulty of the work was increased all along from the relatively easy initial levels by making the worksheet exercises more complex at a pace suitable to the learners' progress. This format is given in Table 12. "Instructors" refers to the sJsK-trained community members.

Because of the very low levels of literacy encountered, interventions to Grade 4 and Grade 7 learners followed the same format, as can be seen from Table 13. The complexity level of the reading and worksheet exercises differed between the grades to ensure that the work matched the ability of the learners.

**Table 11: Hours of normal and intervention tuition received by school type and grade during the second semester**

Grade	Control schools: Normal tuition Hours	Project schools: Normal tuition Hours	Project Intervention Hours	Intervention %
1	36-54	24-42	12	22,2%
4	90	78	12	13,3%
7	72	48	24	33,3%

**Table 12: Standard format of one-hour intervention sessions to Grade 1 learners during the second semester**

Item	Time	Content	Delivered by	Materials
1	10 minutes	Song in English	Instructors	None
2	10 minutes	Phonetic alphabet	Instructors	Wall and individual charts
3	10 minutes	Flash card drills	Instructors	Flash cards
4	30 minutes	Worksheet exercises	Instructors	sJsK worksheets in workbooks
<b>Total:</b>	<b>60 minutes</b>			

**Table 13: Standard format of one-hour intervention sessions to Grade 4 and Grade 7 learners during the second semester**

Item	Time	Content	Delivered by	Materials
1	5 minutes	Phonetic alphabet	Instructors	Wall and individual charts
2	15 minutes	Basic phonics drills	Instructors	sJsK worksheets in workbooks
3	15 minutes	Reading in Xitsonga and English	Instructors	sJsK bilingual reading booklets
4	25 minutes	Worksheet exercises based on memory training, comprehension, literacy and mental disciplines	Instructors	sJsK worksheets in workbooks
<b>Total:</b>	<b>60 minutes</b>			

The selection, numbers, training and facilitation (i.e., intervention) work undertaken by the instructors recruited from the community for the task and of teachers are now further elaborated on and synthesised.

### *Instructor selection and training*

Instructors were selected and trained in two tranches. There were 104 applicants from the school communities in the first tranche and 65 in the second tranche. These applicants were shortlisted into groups of 50 (first tranche) and 30 (second tranche), before second interviews and final selections were made.

In total 25 Instructors were selected from the first tranche and 15 from the second tranche. Some of these successful applicants dropped out either during training or in the early phases. Forty Instructors initially started the programme. Eight dropped out over the year and were not replaced.

These community members had either a matriculation certificate or an old Standard 8 qualification (now Grade 10) as their highest achievement. All had to be reasonably fluent in both written and spoken English. They had no formal training in education other than the training received as part of the intervention.

All Instructors selected were tested using the Schonell's Reading test and all did a Grade 8 comprehension test as part of the final selection procedure.

The Instructors received three weeks initial training on practical methodology and pedagogic practices and received an additional two hours per week of content-related training to ensure that they were on top of the material they needed to present. Ongoing instructor training for community members was carried out at a rate of twelve two-hour sessions (one two-hour session per week) over the period to ensure that the Instructors were familiar with the work they were to present.

The Instructors also received two weeks training to familiarise them with the materials to be used in the assessment and in conducting and marking of the assessments.

The methodology and pedagogic training developed their presentation ability and classroom skills, which in turn enabled them to deliver the lessons according to the designed structure and format.

The content-related training ensured that the Instructors were familiar with the work they would be presenting in the week ahead, and that they were comfortable with both the presentation and support aspects required to successfully deliver the work to the learners.

### *Teacher training*

The teacher training was implemented in the second semester of 2013. However, because of a combination of factors including the South African Annual National Assessments, local civil unrest and CAPS training, only six one-hour interventions were delivered per school during this period. Attendance by teachers varied, but was in general low due to the aforementioned factors. The training sessions were open to all teachers of all subjects at the schools as literacy is a thread common to all subjects.

The teacher training was carried out by two retired teachers (Trainers) who had been recruited to run fortnightly sessions at each school in the afternoons after school. Each session offered practical suggestions for the teachers to use in their teaching, combined with follow-up discussions.



The sessions offered coping strategies for large classes with huge learner backlog issues. Teacher attendance in the training sessions was sub optimal, largely as a result of other departmental demands on the teachers such as CAPS training.

*The intervention content*

The learner-focused English Literacy intervention is a specially designed literacy programme which:

- Tests every learner to assess where his/her literacy level is – including phonetic knowledge (phoneme, grapheme, etc.), word-attack and word-recognition skills, comprehension and written skills as well as reading ability.
- Has specially designed materials which rebuild missing foundations and rapidly bring the learner up to the literacy levels required to engage successfully with the curriculum.
- Is delivered in the classroom during contact time by trained community members.

The intervention focuses on rebuilding learner foundations in English Literacy by supplying and supporting a specifically designed learner-centred learning experience. To achieve this, the intervention interacts with the learners for either one hour (Grades 1 and 4) or two hours per week (Grade 7). This results in a total contact time of 18 hours for Grade 1 and 4 learners and 30 hours for Grade 7 learners. The intervention replaces one FAL hour per week for Grade 1 and 4, and two FAL hours per week for Grade 7 (after June – prior to June the Grade 7s received one hour per week).

The intervention was delivered entirely by the 32 community member Instructors. During the contact time, community members who have been selected and trained as Instructors deliver the entire intervention lesson to the learners. The class is divided into groups of approximately eight to ten learners, and an Instructor is assigned to each group.

Each learner is supplied with a structured workbook and reading book appropriate to his or her ability. The Instructors guide the learners through the assigned lesson and mark and correct learners’ work in real time. At a class level, resources such as flash-card sets and wall charts are supplied, specific to the lesson being delivered.

Each learner received six workbooks during the intervention, varying between 14 and 28 pages each. The Grade 4 learners read six dual-medium (English / XiTsonga) graded reading books, each book being read at least four times in each language, making a total of 48 book readings. The Grade 7 learners read 12 reading books, each book also being read at least four times in each language, making a total of 96 book readings.

One of the key items of interest is the difference in human/adult contact time (per learner) between the control and project schools. Table 14 shows the learner contact time per group for the 2013 year. These figures were calculated by taking the number of teaching person hours available in the respective control and project groups and dividing it by the number of learners in each respective group.

**Table 14: Comparison of learner contact time between control and project schools**

Contact hours per learner in -	Control Teacher	Project school situation			Project vs control
		Teacher	Instructor	Total	
Grade 1	1,4	1,0	1,5	2,5	174%
Grade 4	2,5	1,9	1,8	3,7	150%
Grade 7	2,3	1,4	3,6	5,0	215%

As Table 14 shows, project-school Grade 1 learners received 174% of the human contact time that their peers in the control schools got. The project-school Grade 4 learners received 150% of the human contact time, and the project-school Grade 7 learners received more than double the contact time (215%) of their control-school peers. The project-school learners can be said to have received up to more than double the amount of attention that their control-school peers received.

Project-school learners receive FAL tuition from their teachers as well as from the Instructors, and it is worth noting that the project-school learners received almost as much attention (Grade 4 – 2,5 for control vs 1,8 for project) to considerably more attention (Grade 7 – 2,3 for control vs 3,6 for project) from the Instructors in the respective 18 and 30 hours of intervention than the control-school learners received during their entire school year.

In terms of costs, the total average cost per learner was R795,73 per learner per year, calculated by taking the total cost of the intervention and dividing it across the total number of project-school learners.

The evaluation design and consistency of intervention volumes across schools do not allow for isolating any effects for sub-components of the intervention. The intervention is therefore treated as a full package.

The discussion of impact is separated into sub-sections covering the outcomes at Grade 1, 4 and 7 levels because a different set of tests was administered at each level. Test contents are introduced as relevant below. They are not detailed beyond basic descriptions because the intervention reports compiled by sJsk, especially their first main report after Semester 1, comprised that information.

A few final observations delineate the first full-year impact analysis further. The assessment instruments were selected with the direct intervention benefits in mind. This was partly motivated by avoiding expected floor effects should more complicated measures be used early on. Instrument or criterion-measure selection had also been based on the expectation that the literacy and language gaps to be corrected, that is, the regeneration of lost conceptual understanding that had to take place, and the immediate proficiencies related to “code” (decoding) that needed to be rebuilt, would be the initial thrust. This means that the intervention and evaluation teams both accept that improved understanding and production of “meaning” by learners are the eventual successes aimed at. However, long lead and lag times in the regular schooling trajectory are not foreign concepts. Therefore, the early foundational successes have been expected to be dramatic in terms of repairing and replacing basic literacy and language building blocks. Once these learner tools are sharpened and their toolkit fitted out with confidence, subsequent gains in vocabulary, reading fluency and more complex grammatical proficiency and understanding is expected to lead to broader academic growth in relation to increasing achievement both in the language curriculum and subjects as a whole, and in subjects across the curriculum. This is best analysed at a later stage, for instance by using learners’ Annual National Assessment (ANA) results. Such analyses, however, are not expected to show the effect of regenerated foundations after only one year.<sup>16</sup>

### 3.2.1 Grade 1

The four tests administered at this level comprised three sub-tests from the Early Grade Reading Assessment (EGRA), as well as an Oral Pictorial Test. As a result, it entailed oral administration of the following:

---

<sup>16</sup> Some 2012 and 2013 ANA results were obtained during April 2014 for the Grade 4 cohort of 2013. No similar trend results exist for the Grade 1 and Grade 7 cohorts. Data availability was incomplete and erratic across language subjects and schools. Exploratory calculations seem to confirm that no effects would be observed in these results yet.

- EGRA – Letter-Sound sub-test (recognising letters - “vowels” and “consonants” – in a phonetic way from the alphabet)
- EGRA – Word Recognition sub-test (recognising known words)
- EGRA – Non-Word Decoding sub-test (ability to read unfamiliar stimuli out loud)
- Oral Pictorial test (naming 10 basic objects, animals or children from culturally appropriate sketches).

It has to be acknowledged that Grade 1 learners, especially at the baseline, have had a very limited exposure to any formal assessment and the school world at large. Hence there are limits to what they can be exposed to in Grade 1. This would apply particularly to written text and writing.

The total sample for which paired scores were available ranged from 698 to 701 learners depending on the sub-test. Of these learners, 252 or 253 were from control schools and between 445 and 449 from project schools. There was minimal variation of this respondent number between sub-tests.

The effect of basing reporting only on paired scores was investigated before going through with such reporting. The score distributions of learners whose data belonged to either a paired-score or unpaired-score sub-group were analysed and are reported in Annexure 1. It can be seen from the mean achievement scores and their standard errors and standard deviations, along with indications of statistical probability, that these two sub-groups do not differ much. In fact, scores from one sub-test (EGRA Letter Sound) would account for four of the five flagged and possible attrition effects, had the 10%-level of probability been considered. In addition, achievement gains from February to November never revealed different trajectories when the paired and non-paired groups are compared, thus suggesting that fortuitous or random sub-group differences on the basis of other unrelated factors may have been the case. What is more, although these probability values could to some degree underpin conclusions about statistical significance, it is advisable not to rely on them in the present report, as explained earlier. The main reasons are that computing technical corrections on the basis of clustering of learners into schools (and classes) would diminish, if not remove fully, the chances of finding significant effects. It therefore makes more sense in this evaluation report to look at the practical meaning of any differences and reserve further sophisticated analyses for publications in academic journals in future.

The resulting mean percentage scores from the paired comparisons, the various percentage-point and percentage changes and the difference-in-difference scores were as reflected in Table 15. Standard deviations are reported to enable interpretation of differences in gain over the baseline. Although statistical significance outcomes are reported, their interpretation should factor in the previously explained emphasis of this report on practical rather than technical interpretation. The results signify that there were no (statistical) differences between learners at the outset, that there were (statistical) differences after one year, and that there were strong gains for project-school learners in all four sub-tests.

Gains in the word-based sub-tests take longer to manifest than proficiencies in the other two sub-tests. This is evident from the scores for learners from project and control schools after one year. However, the erratic patterns that were still observed by June 2013 have largely disappeared. Nevertheless, percentage-change interpretations will be meaningless in the case of near-zero baseline scores. The latter essentially reflect the (expected) absence of the proficiencies concerned at the beginning of Grade 1. In relation to sub-tests linked to decoding proficiency, which would signal the expert regeneration inputs of the interventions, project-school learners have “out-gained” learners from control schools more significantly.

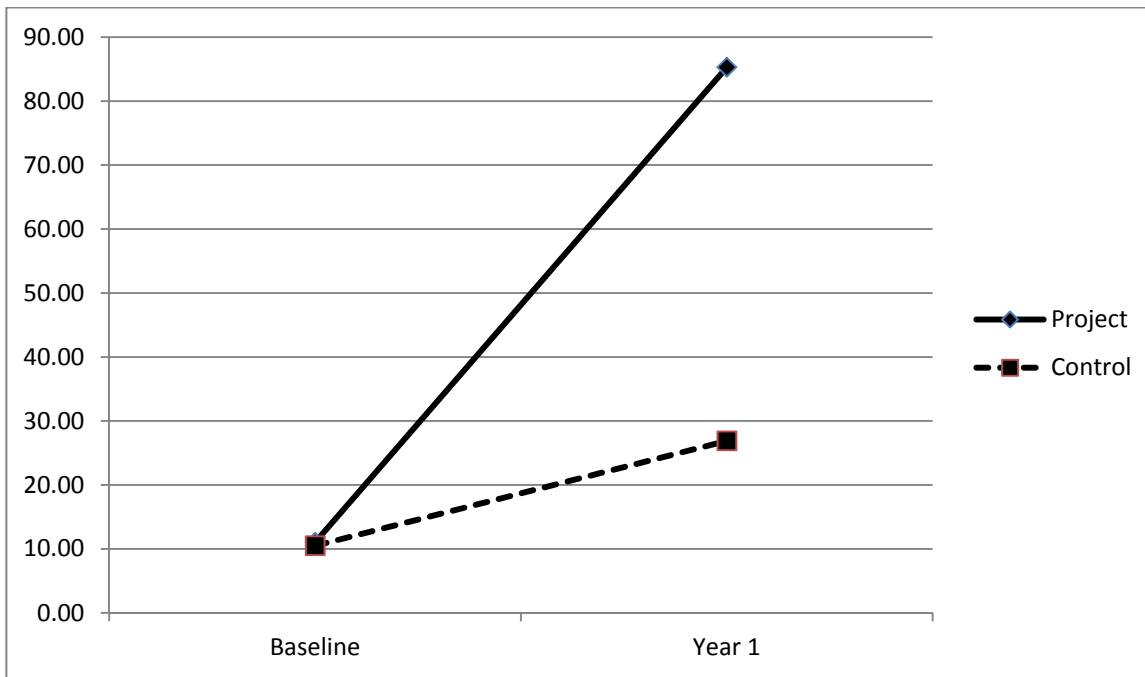
The four outcomes are depicted graphically in Figures 3 to 6. Where zero baselines do not render calculating improvement percentages meaningless, these findings represent improvements of 675% and 563% on average for learners in project schools compared to improvements of 156% and 337% for learners in the control group respectively on the Letter Sound and Oral Pictorial sub-tests. Although low baselines still cause large percentage-gain figures, the comparative picture remains realistic and meaningful. As required, learners after a year display substantive proficiency off a low base. These outcomes are very close to the ones found in relation to all the sub-tests by JSK and reported in their annual report.

**Table 15: Mean percentage achievement, percentage point change, difference-in-difference and percentage change figures for Grade 1 learners by school type**

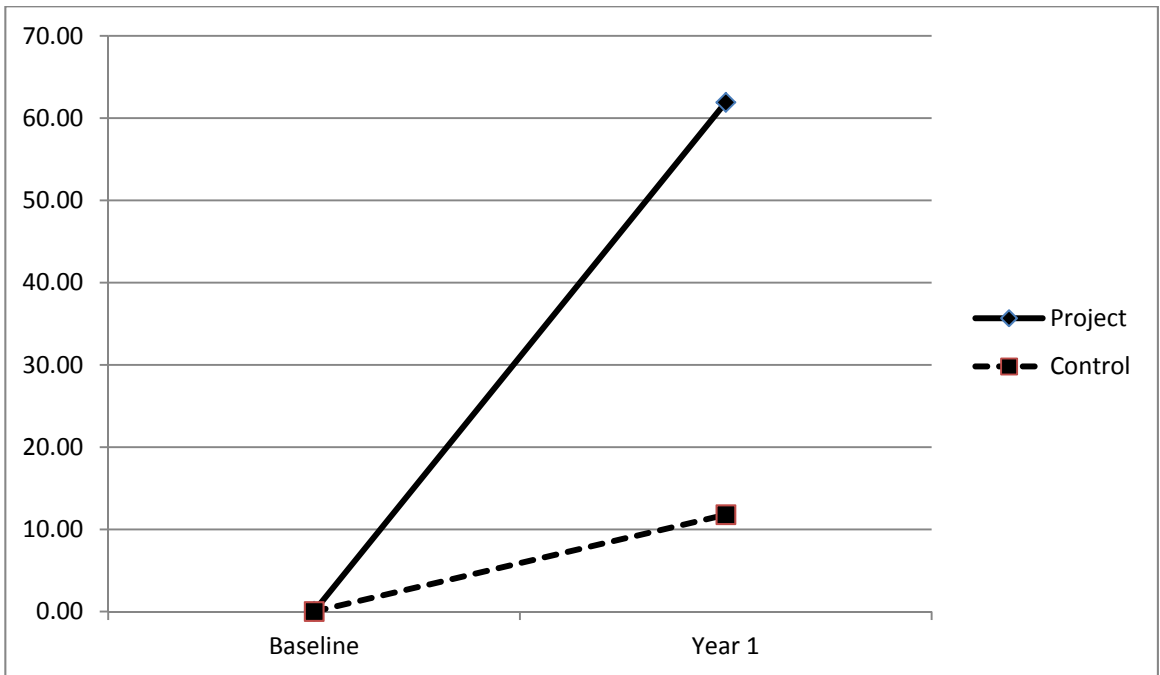
Test	School subset	Baseline % (& SD)	End Year 1 % (& SD)	Change % pts (& SD)	Diff-in-diff % pts (& SD)	Change % (& SD)
EGRA Letter Sound	Project	11,0 (18,5)	***85,3 (18,0)	***74,3 (24,0)	57,9 (3,0)	674,9 (4,0)
	Control	10,5 (15,9)	26,9 (21,6)	16,4 (21,0)		156,0 (1,0)
EGRA Word Recognition	Project	0,0 (0,2)	***61,9 (24,6)	***61,9 (24,6)	50,1 (@)	@ (@)
	Control	0,0 (0,1)	11,8 (19,0)	11,8 (18,9)		@ (@)
EGRA Non-Word Decoding	Project	0,2 (4,7)	***73,3 (23,2)	***73,1 (23,7)	55,0 (@)	@ (@)
	Control	0,0 (0,0)	18,1 (26,2)	18,1 (26,2)		@ (@)
Oral Pictorial	Project	13,6 (13,8)	***90,1 (13,8)	***76,5 (20,2)	33,8 (2,2)	563,1 (5,5)
	Control	12,7 (12,9)	55,3 (27,4)	42,7 (28,6)		337,0 (3,3)

\* p<0,05    \*\*p<0,01    \*\*\*p<0,001    (No corrections calculated for this report for effect of cluster sampling)  
 @ Low baseline (non-existing early proficiency) renders these figures meaningless.

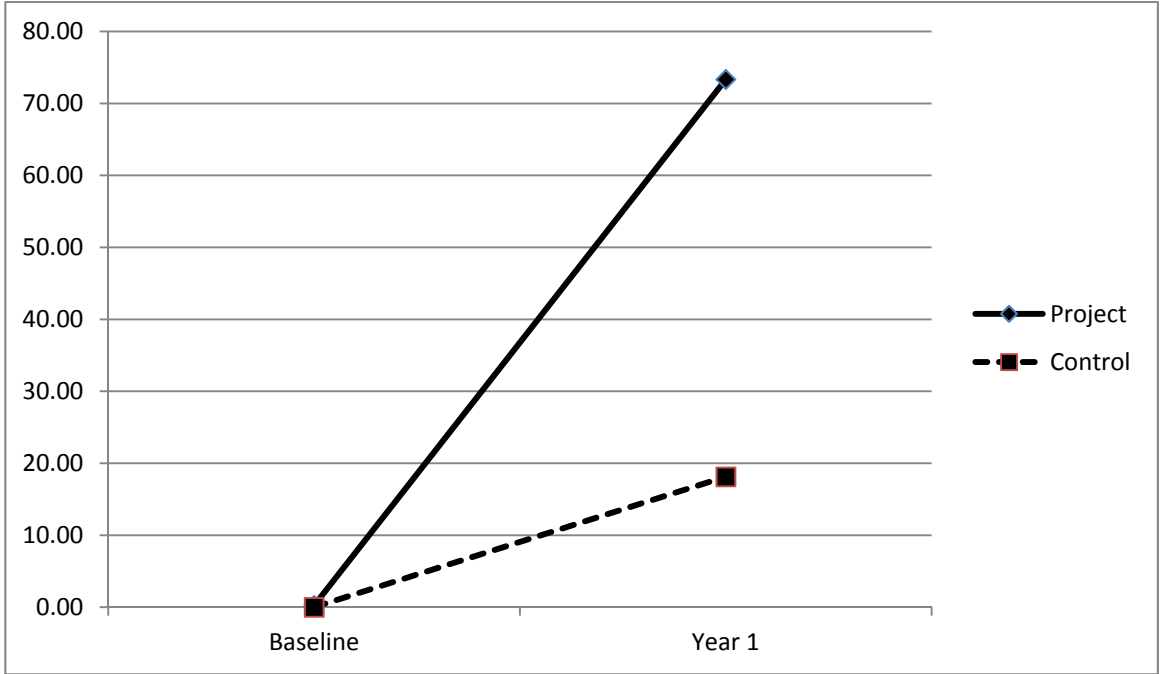
Note: To explore the extent of the effect of intervention status (project vis-à-vis control school), a basic value-added regression model was run. It regressed learners' November scores on their baseline scores (February) and intervention status. As explained before, the intention is to explore such quantitative, technical and statistical matters further in future, preferably closer to the final evaluation and in academic outputs. Hence, no additional tables are included. Some core statistics appear in an endnote<sup>1</sup>.



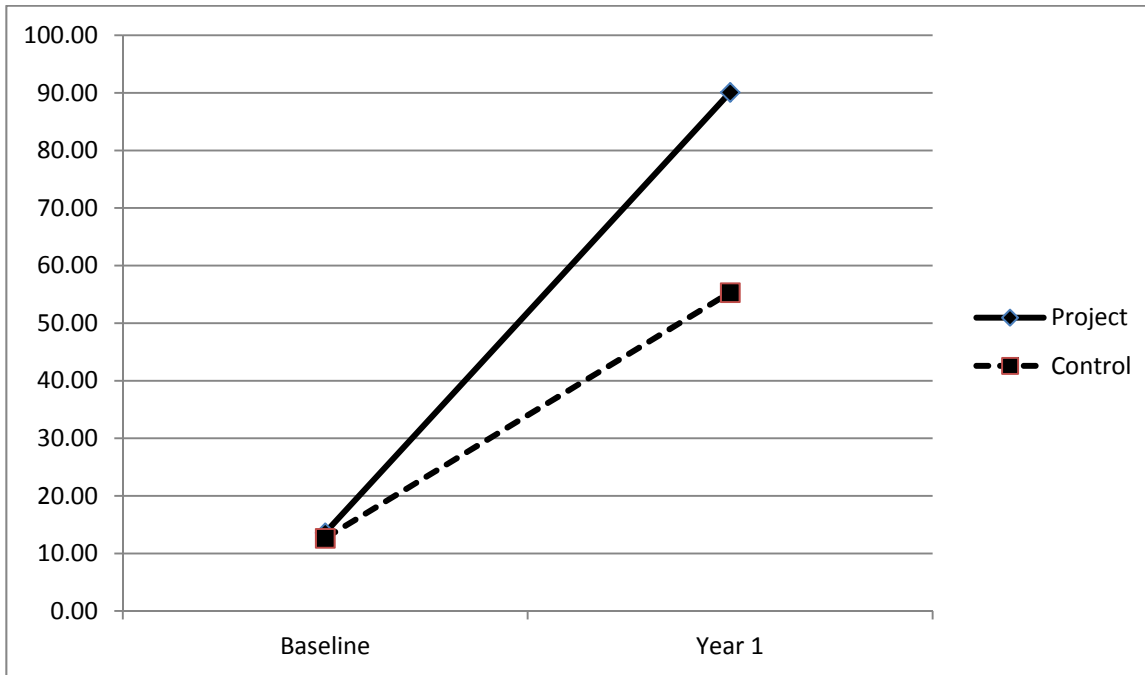
**Figure 3: Grade 1 mean learner scores (%) on the EGRA Letter Sound sub-test over time by school type**



**Figure 4: Grade 1 mean learner scores (%) on the EGRA Word Recognition sub-test over time by school type**



**Figure 5: Grade 1 mean learner scores (%) on the EGRA Non-Word Decoding sub-test over time by school type**



**Figure 6: Grade 1 mean learner scores (%) on the Oral Pictorial test over time by school type**

For improved alignment with school-level interpretation needs, learner achievement increases are also compared briefly in terms of the percentage of learners over time, for project and control schools, who “mastered 50% of their work” on the basis of the assessments. The outcome is reflected in Table 16. The results signify that there had been no (statistical) differences between learners at the outset, but that by June, with the exception of Non-Word Decoding, (statistically) different achievement between children from project and control schools had occurred. In addition, it also signifies strong gains for project-school learners in all four sub-tests by November.

**Table 16: Percentage of learners in project and control schools achieving a mark of at least 50% in the four Grade 1 sub-tests during testing in February, June and November 2013**

Test (n)	School subset	Baseline	June	November
EGRA Letter Sound (n=700)	Project	6,7	***36,1	***95,7
	Control	4,7	12,6	17,0
EGRA Word Recognition (n=698)	Project	0,0	**3,6	***72,8
	Control	0,0	0,0	6,3
EGRA Non-Word Decoding (n=699)	Project	0,2	6,1	***87,0
	Control	0,0	4,2	16,3
Oral Pictorial (n=701)	Project	2,4	***56,3	***97,8
	Control	3,2	18,1	58,7

\* p<0,05    \*\*p<0,01    \*\*\*p<0,001    (No corrections calculated for this report for effect of cluster sampling)  
 Note: p-values based on Phi and Cramer’s statistics

A final note is made to point out how low the achievement level is that learners from a small selection of regular public schools achieve after a full year of schooling. They would be the learners from the control schools. Their situation would imply the usual conditions and outcomes of teaching and learning in schools under government (provincial education department) jurisdiction. The specific outcomes of concern are learners’ achievement in the suite of EGRA tests, which could be considered as providing a useful international benchmark. The two specific cases in point are learners’ average scores of below 30% on these tests, and the fact that fewer than 20% of the learners reached an acceptable achievement level by the end of Grade 1. Tables 15 and 16 reflect the necessary findings.

### 3.2.2 Grade 4

Seven tests were administered at this level and eight test scores were used in the comparisons. Besides the Schonell Reading Age assessment, three sub-tests from EGRA, and three Literacy sub-tests with their total score comprised the assessment instruments / scores. The complete list of administered tests entails:

- Schonell Reading Age assessment (columns of progressively unfamiliar and difficult meaningful words / vocabulary)
- EGRA – Letter-Sound sub-test (recognising letters - “vowels” and “consonants” - in a phonetic way from the alphabet)
- EGRA – Word Recognition sub-test (recognising known words)
- EGRA – Non-Word Decoding sub-test (ability to read unfamiliar stimuli out loud)
- Literacy 1 sub-test (multiple-choice questions (MCQ) based on a reading passage; cf 4<sup>th</sup> sub-test)
- Literacy 2 sub-test (sentence ordering exercise)
- Literacy 4 sub-test (comprehension questions based on a reading passage comprising a folklore story about a rabbit, with open-ended responses)
- Literacy Total score (the average percentage based on the sum of the items in the previous three, out of a maximum of 15).

The total sample for which paired scores were available ranged from 847 to 854 for the Schonell Reading Age and EGRA tests and from 878 to 883 for the Literacy test scores. The exact numbers depend on the sub-test. Of these learners, 499 to 502 for the Schonell Reading Age and EGRA tests and 526 or 527 for the Literacy test scores were from project schools. The numbers were 347 to 352 for the Schonell Reading Age and EGRA tests and from 351 to 356 for the Literacy test scores for control-school learners. Within these two clusters of tests, there was minimal variation of these respondent numbers between sub-tests. These outcomes were determined by learner attendance patterns on the days of testing.

The effect of attrition in securing paired-score datasets was again investigated before going through with the reporting below. The score distributions of learners whose data belonged to either a paired-score or unpaired-score sub-group were analysed and are reported in Annexure 2. The mean achievement scores and their standard errors and standard deviations, along with indications of statistical probability, indicate that these two sub-groups do not differ substantively. In fact, only one set of scores (EGRA Letter Sound) account for a difference within the 1%-level of probability. Five other sets, all within the 5%-level of probability, were flagged for possible attrition effects. However, these sets were all distributed inconsistently across, project and control groups, stage of the year of testing, and sub-test at stake, further diminishing the chances of systematic bias while pointing to rather random influences on the basis of unrelated factors. As before, it would not be prudent to rely on the statistical route too much for making final interpretations given the fact that corrections for clustered sampling have not been made (yet), which would greatly diminish statistically significant outcomes. The practical indications discussed so far, as also reflected in the annexure, provide enough support for the purposes of this report to conclude in favour of the absence of bias.

The resulting mean percentage scores (with the exception of the results from the Schonell test which gives a reading age estimate) are reflected in Table 17 below. Standard deviations are reported to enable interpretation of differences in gain over the baseline. Although statistical significance outcomes are reported, their interpretation should factor in the previously explained emphasis of this report on practical rather than technical interpretation. The results signify that there were some statistical differences between learners from the two groups already at the outset, and that statistical differences by year-end had generally swung in favour of the project group. It is further revealed that statistically significant gain differences occurred over one year for project-school learners in seven of the eight tests.

The significantly higher gains for learners from project schools compared to those from control schools in the Schonell Reading Age, three EGRA, Literacy 2 (sentence ordering) and Literacy 4 (comprehension) sub-tests and the overall literacy score were, are depicted graphically in Figures 7 to 14 further down. These findings again are very similar to those reported by sJsK, although the HSRC’s presentation kept the EGRA

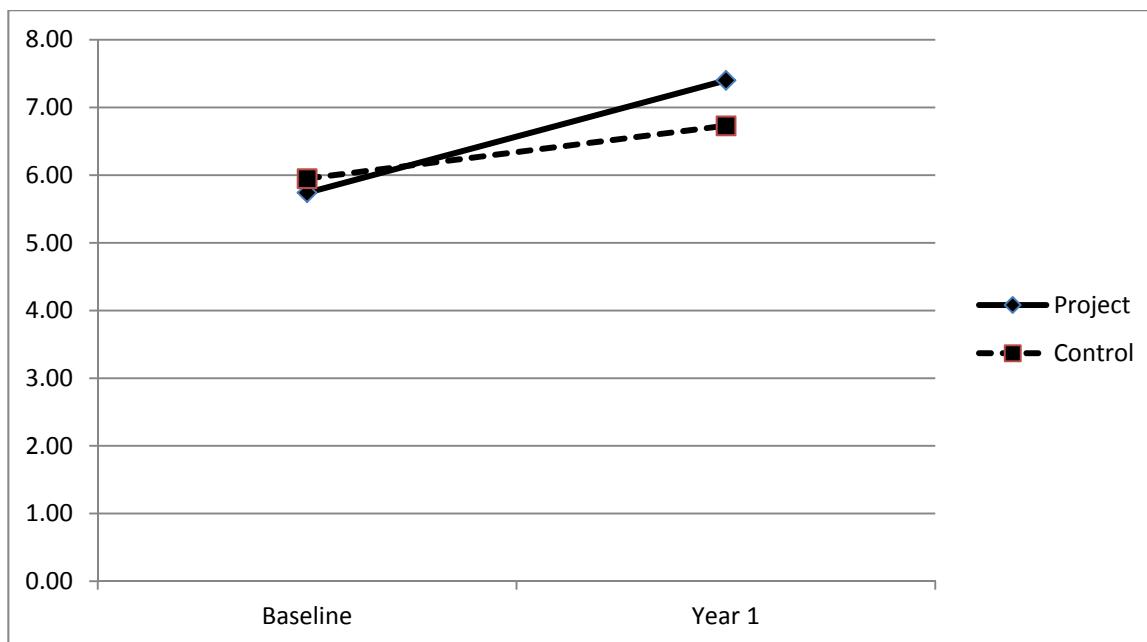
sub-tests separate during analyses. Achievement gains pertaining to the three EGRA tests were particularly strong. These findings represent improvements ranging from almost 30% to as high as 268% and over 400% for learners in project schools compared to changes from just more than 10% to almost 150% for learners in the control group. It is reiterated that Reading Age scores are not measured in percentages, but in years.

**Table 17: Mean percentage achievement (age estimate for Schonell test), percentage point change, difference-in-difference and percentage change figures for Grade 4 learners by school type**

Test	School subset	Baseline % (& SD)	End Year 1 % (& SD)	Change % pts (& SD)	Diff-in-diff % pts (& SD)	Change % (& SD)
<b>Schonell Reading Age (not %)</b>	<b>Project</b>	5,7 (1,6)	***7,4 (1,4)	***1,7 (1,6)	0,9 (0,4)	28,9 (1,1)
	<b>Control</b>	5,9 (1,3)	6,7 (1,0)	0,8 (1,2)		13,2 (0,6)
<b>EGRA Letter Sound</b>	<b>Project</b>	***25,0 (28,4)	***92,3 (13,6)	***67,2 (30,5)	61,4 (2,0)	268,4 (2,4)
	<b>Control</b>	11,9 (14,2)	17,7 (19,7)	5,8 (23,4)		48,8 (0,4)
<b>EGRA Word Recognition</b>	<b>Project</b>	38,3 (32,2)	***85,2 (19,5)	***46,8 (27,3)	28,1 (0,9)	122,2 (1,5)
	<b>Control</b>	***46,6 (31,0)	65,4 (28,0)	18,8 (19,9)		40,2 (0,6)
<b>EGRA Non-Word Decoding</b>	<b>Project</b>	39,4 (32,2)	***88,9 (17,7)	***49,4 (30,8)	34,4 (1,0)	125,4 (1,5)
	<b>Control</b>	***48,5 (30,0)	63,5 (25,1)	15,0 (24,7)		31,0 (0,5)
<b>Literacy 1 MCQ</b>	<b>Project</b>	***21,5 (25,4)	***39,3 (31,7)	17,8 (38,2)	2,2 (0,0)	82,9 (0,7)
	<b>Control</b>	12,5 (23,4)	28,1 (29,3)	15,6 (34,4)		124,8 (0,7)
<b>Literacy 2 Sentence Order</b>	<b>Project</b>	20,2 (24,0)	***38,3 (31,8)	***18,1 (39,0)	13,8 (0,6)	89,3 (0,8)
	<b>Control</b>	23,4 (23,4)	27,7 (27,2)	4,3 (35,9)		18,4 (0,2)
<b>Literacy 4 Open-ended</b>	<b>Project</b>	2,8 (7,2)	***14,7 (15,6)	***11,9 (16,0)	7,0 (1,1)	426,0 (1,7)
	<b>Control</b>	3,4 (8,5)	8,4 (16,7)	5,0 (15,1)		144,0 (0,6)
<b>Literacy Total</b>	<b>Project</b>	11,2 (9,6)	***25,9 (17,9)	***14,7 (19,1)	7,8 (0,9)	131,0 (1,5)
	<b>Control</b>	10,5 (10,4)	17,4 (16,8)	6,8 (16,3)		64,9 (0,7)

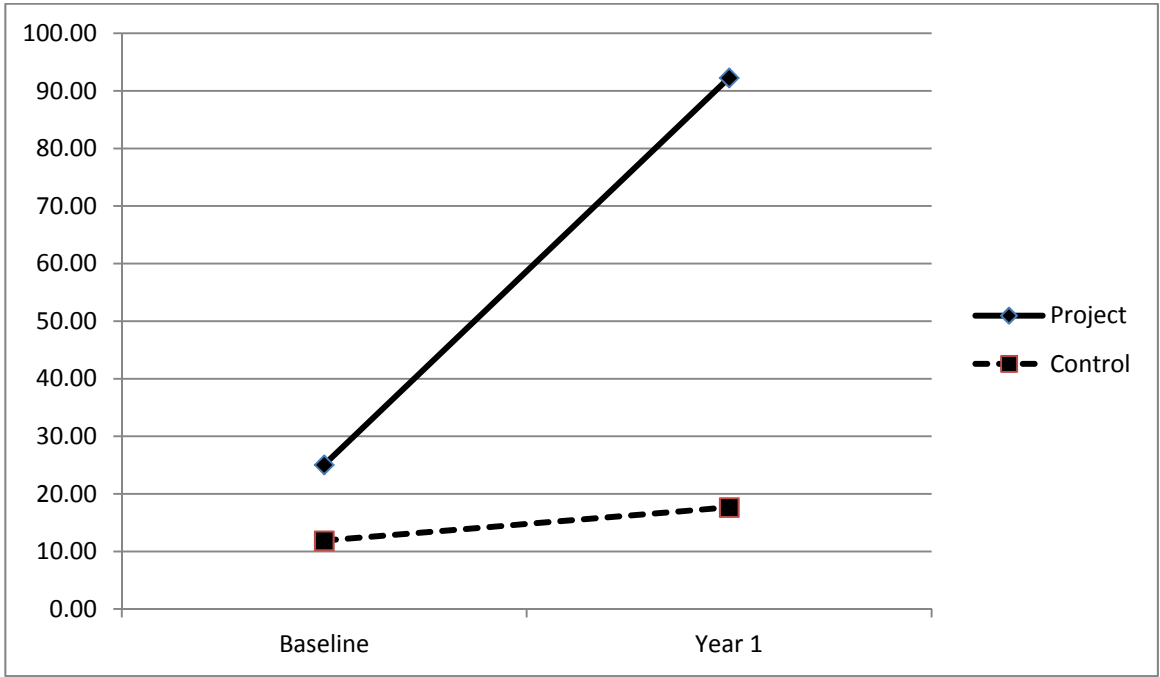
\* p<0,05    \*\*p<0,01    \*\*\*p<0,001    (Correcting p-values after cluster sampling has not been done)

**Note:** To explore the extent of the effect of intervention status (project vis-à-vis control school), a basic value-added regression model was run. It regressed learners' November scores on their baseline scores (February) and intervention status. As explained before, the intention is to explore such quantitative, technical and statistical matters further in future, preferably closer to the final evaluation and in academic outputs. Hence, no additional tables are included. Some core statistics appear in an endnote<sup>ii</sup>.

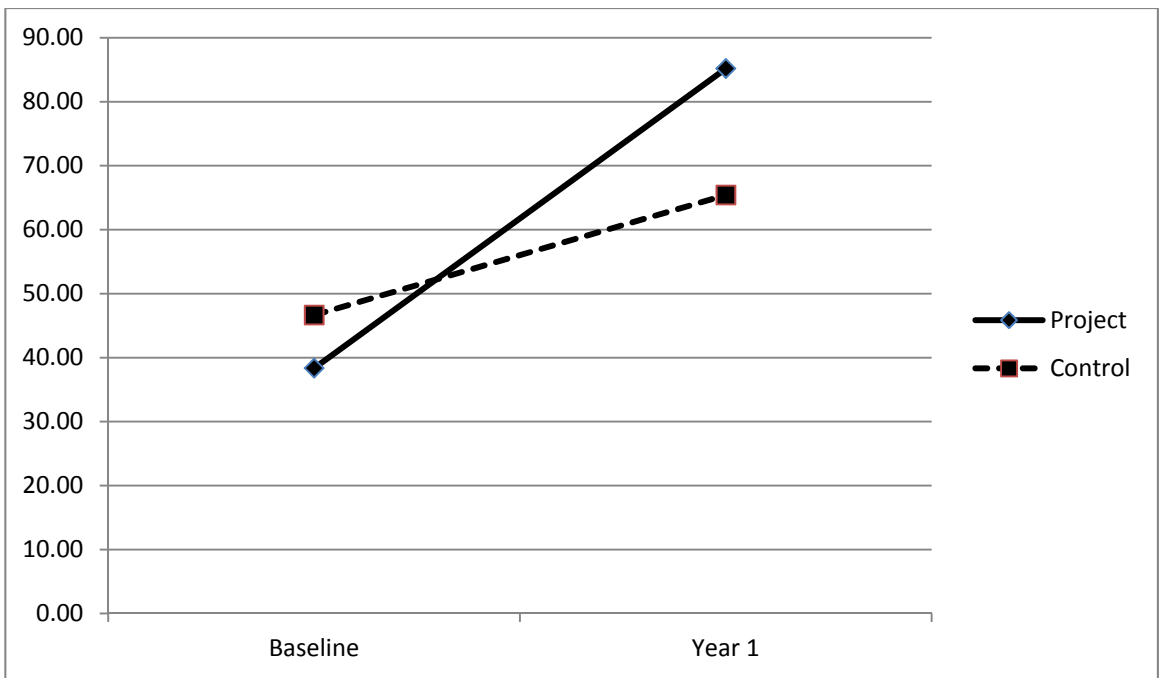


**Figure 7: Grade 4 learner mean scores (reading age) on the Schonell Reading Age test over time by school type**

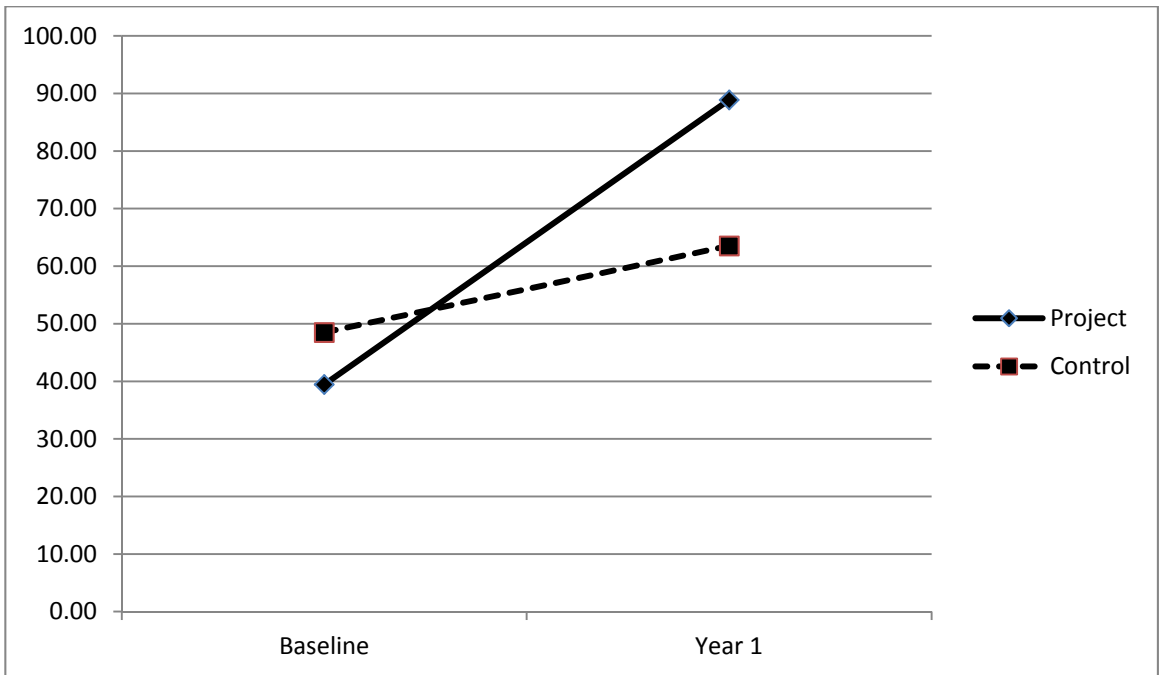




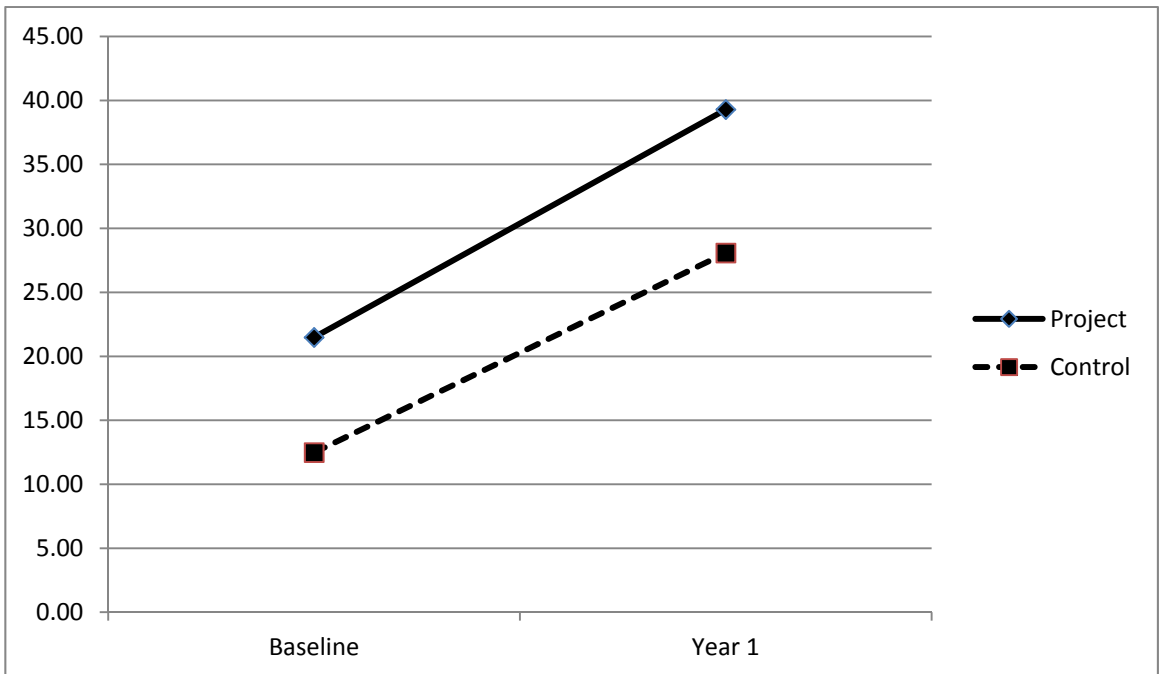
**Figure 8: Grade 4 mean learner scores (%) on the EGRA Letter Sound sub-test over time by school type**



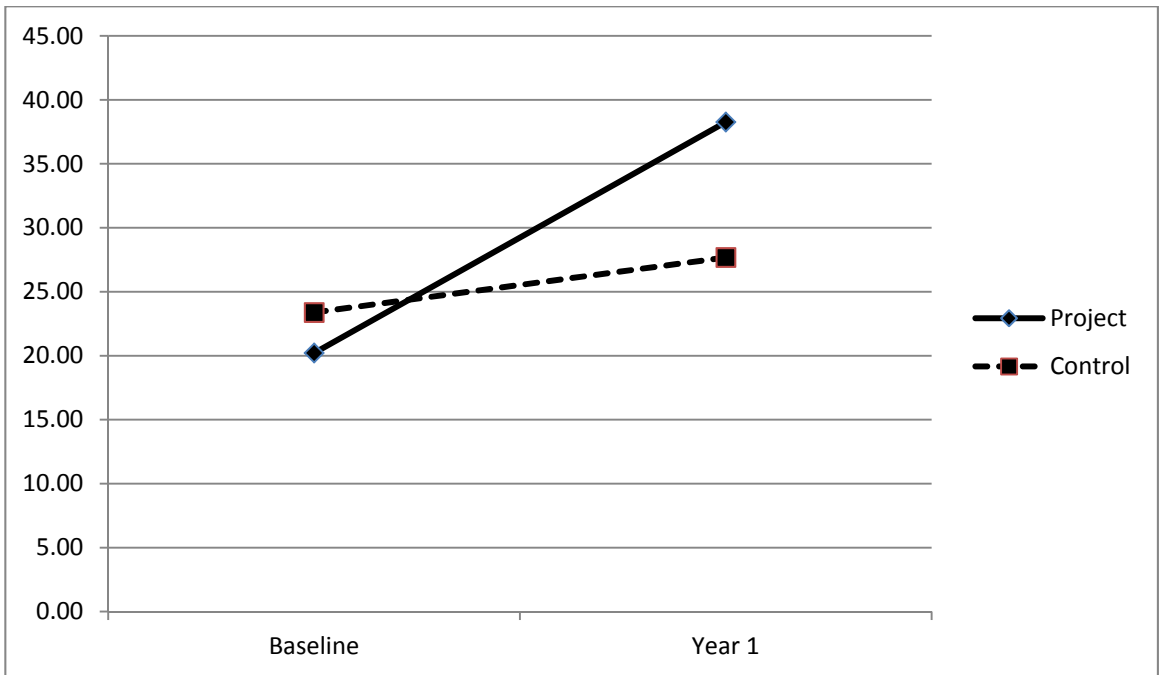
**Figure 9: Grade 4 learner mean scores (%) on the EGRA Word Recognition sub-test over time by school type**



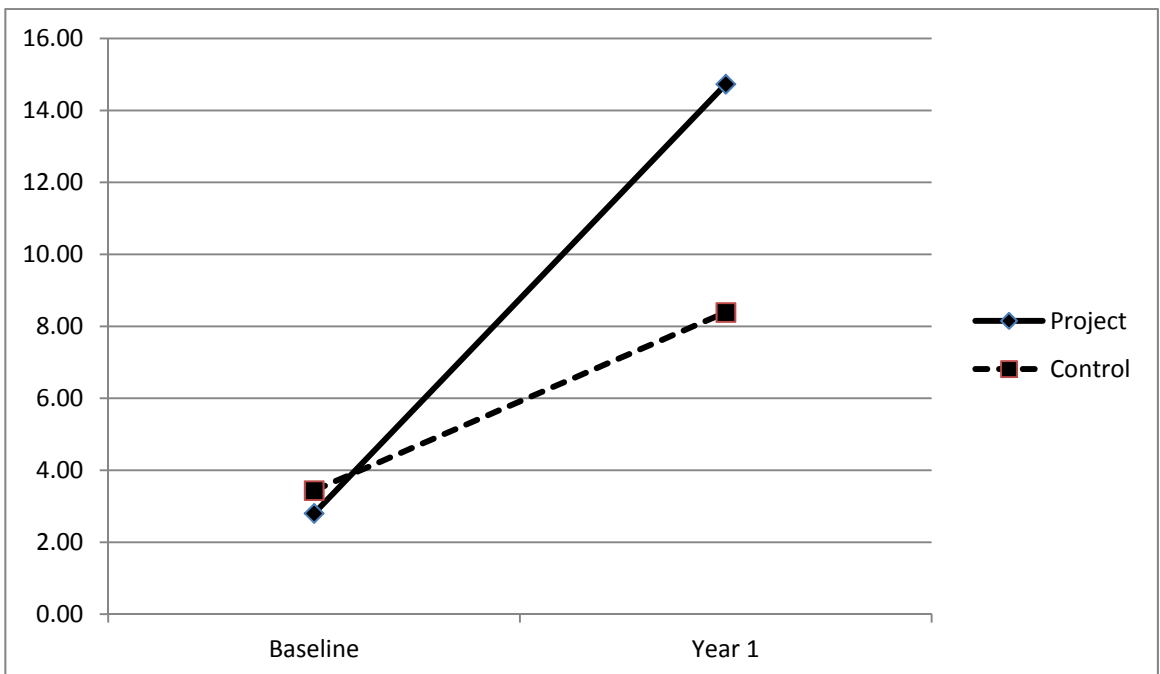
**Figure 10: Grade 4 learner mean scores (%) on the EGRA Non-Word Decoding sub-test over time by school type**



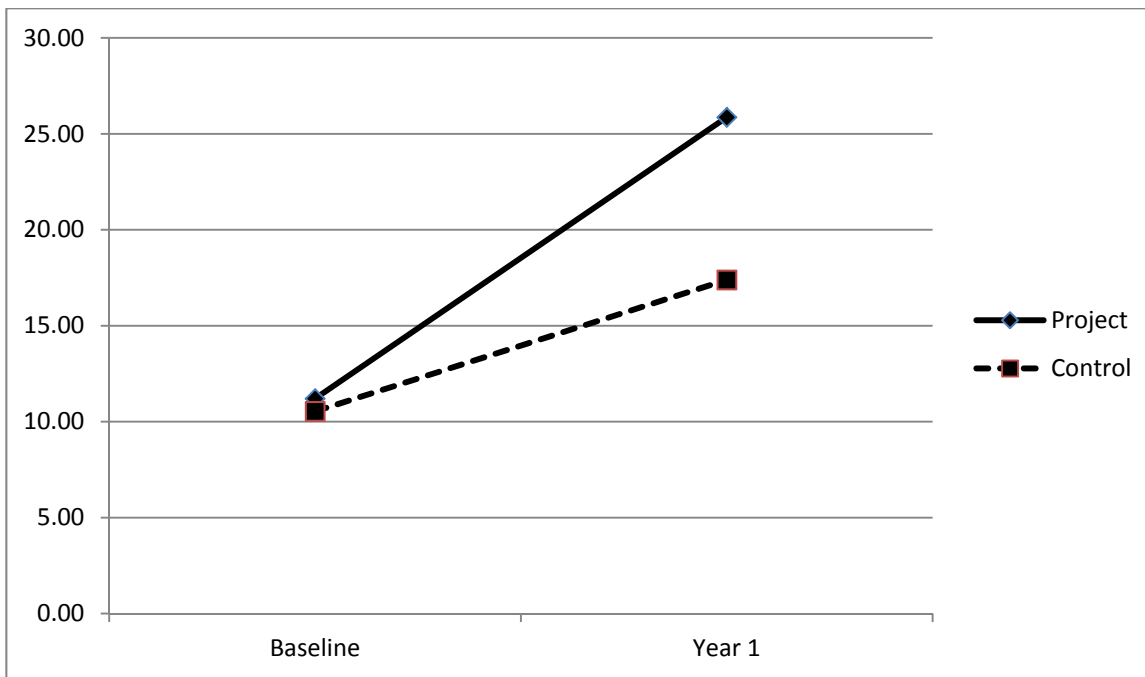
**Figure 11: Grade 4 learner mean scores (%) on the Literacy 1 (MCQ) sub-test over time by school type (note the reduced %-scale on the y-axis)**



**Figure 12: Grade 4 learner mean scores (%) on the Literacy 2 (Sentence Ordering) sub-test over time by school type (note the reduced %-scale on the y-axis)**



**Figure 13: Grade 4 learner mean scores (%) on the Literacy 4 (Open-ended items) sub-test over time by school type (note the small scale on the y-axis)**



**Figure 14: Grade 4 learner mean scores (%) on the overall (aggregated) Literacy test over time by school type (note the small scale on the y-axis)**

For improved alignment with school-level interpretation needs, learner achievement increases are briefly compared further on the basis of the assessments reported on above. This is done now in terms of the percentage of learners over time, for project and control schools, who “mastered 50% of their work” or achieved the desired reading age. The outcome is reflected in Table 18.

It is evident from Table 18 that after one year a significantly higher proportion of learners in project schools achieved a mark of at least 50% on all but one test compared to control-school learners. However, one also has to look at the trend of increase in these proportions of learners over time, as the baseline situation in February inconsistently favoured either the project or control groups, depending on the test. The situation nevertheless had swung in favour of project-school learners by June, with the non-significant exceptions of the reading age and overall literacy scores. In summary, the findings signify that there were some statistical differences between learners at the outset. Sometimes learners from project schools were favoured, but more often not. By mid-year, for the three EGRA sub-tests, learners from project schools showed statistically different higher achievement than children from control schools. After one year, a much higher achievement outcome was evident for project-school learners in all scores but one (Literacy 4 – Open-ended Comprehension items).

It is also clear, as expected, that proficiencies within the direct aim of the intervention programme, i.e., phonemic and decoding skills, benefitted substantively, while derived competencies, such as comprehension, would follow after a time lag. It would be interesting to follow henceforth when comprehension, reading fluency and age, and related achievement in subjects across the curriculum, would follow. This will be one of the foci during Year 2 (2014). It would thus make sense by then or into 2015 to compare the two sets of learners in relation to external criterion scores such as marks from the Annual National Assessments.

**Table 18: Percentage of learners in project and control schools achieving a reading age of 8 or higher or at least 50% in the seven Grade 4 sub-tests during testing in February, June and November 2013**

Test (n)	School subset	Baseline	June	November
Schonell Reading Age of 8 yrs (n=847)	Project	1,2	2,1	***26,0
	Control	0,9	3,1	6,6
EGRA Letter Sound (n=854)	Project	***21,7	***75,8	***97,8
	Control	3,7	5,2	9,7
EGRA Word Recognition (n=850)	Project	39,4	*69,1	***92,8
	Control	**49,7	60,6	73,7
EGRA Non-Word Decoding (n=850)	Project	41,3	**75,1	***95,8
	Control	***55,3	66,2	75,5
Literacy 1 MCQ (n=881)	Project	**14,1	24,2	***32,1
	Control	8,2	19,5	19,4
Literacy 2 Sentence Ordering (n=882)	Project	20,7	18,8	***44,6
	Control	*26,5	14,3	30,7
Literacy 4 Open-ended (n=878)	Project	0,4	1,6	6,5
	Control	0,6	2,2	5,1
Literacy Total (n=883)	Project	0,6	0,9	**11,2
	Control	0,8	1,5	5,6

\* p<0,05    \*\*p<0,01    \*\*\*p<0,001    (Correcting p-values after cluster sampling has not been done)

Note: p-values based on Phi and Cramer's statistics

One observation that can be emphasised is that teaching in the public control schools only succeeded in raising the reading age of their learners by 0,8 of a year over the course of a year. In effect those learners therefore fell behind slightly further and are now almost a year and a half, at minimum, behind where they should have been. In contrast, learners from project schools caught up almost an additional year, and are now just over half a year behind where they should have been. Tables 17 and 18 provide the necessary details underpinning this observation.

### 3.2.3 Grade 7

Four tests were administered at this level and five test scores were used in the comparisons. Besides the Schonell Reading Age assessment, three Literacy sub-tests with their total score comprised the assessment instruments / scores. The complete list of administered tests entails:

- Schonell Reading Age assessment (columns of progressively unfamiliar and difficult meaningful words / vocabulary)
- Literacy 1 sub-test (multiple-choice questions (MCQ) based on a reading passage; cf 4<sup>th</sup> sub-test)
- Literacy 3 sub-test (paragraph writing exercise)
- Literacy 4 sub-test (comprehension questions based on a reading passage covering the topical issue of Nkosi Johnson and HIV/AIDS, followed by open-ended items)
- Literacy Total score (the average percentage based on the sum of the items in the previous three, out of a maximum of 20).

The total sample for which paired scores were available comprised 874 learners for the Schonell Reading Age assessment and between 924 and 936 learners for the various Literacy scores. For the Schonell Reading Age score, the numbers were 526 and 348 respectively in relation to learners from project and control schools. The Literacy score ranges were between 581 and 584 and between 341 and 352 respectively for learners from project and control schools. There was again very little variation in sub-test sub-samples. This largely depended on learner attendance on the assessment days.

The effect of attrition in securing paired-score datasets was again investigated before going through with the reporting below. The score distributions of learners whose data belonged to either a paired-score or unpaired-score sub-group were analysed and are reported in Annexure 3. The mean achievement scores and their standard errors and standard deviations, along with indications of statistical probability, indicate that these two sub-groups did not differ substantively. In fact, only four sets of scores differed at a probability level smaller than 5%. Three of these differences pertained to greater achievement at the baseline (February) across different tests among learners for whom paired scores were available in comparison with unpaired score sets. The fact that one of these trends reversed up to November, and that it was the only set that reached a level of probability smaller than 1%, diminish the chances that the differences were due to systematic bias. The remaining instance, significant at a level smaller than 0,1%, was for the November comparison for control-school learners on the Schonell Reading Age assessment. Not being supported or confirmed by similar trends could indicate an isolated case without a systematic cause. Again, steering away from over-reliance on such statistical evidence is advocated because corrections for clustered sampling had not even been made for this report. The nature and spread of the limited observed indications and trends, as reflected in the annexure, do not strongly support concluding in favour of bias.

The resulting mean percentage scores are reflected in Table 19. Standard deviations are reported to enable interpretation of differences in gain over the baseline. Although statistical significance outcomes are reported, their interpretation should factor in the previously explained emphasis of this report on practical rather than technical interpretation. The results show that in all cases learners from the control group started off from higher achievement baselines than the project group. By November, all statistical differences changed to now favour project-school learners. The gain scores further confirm the statistically significant greater intervention benefits to project-school on all five test scores over the period of the first year. Even though the sample had been balanced towards the control-school learners, they have been significantly overtaken by the project group. For the Schonell Reading Age assessment and the first and fourth literacy subtests (MCQ items and open-ended comprehension items) the baseline significances only reached or exceeded the 5% level, but not 1%. Therefore, it is clear that project-school learners gained much more and were without exception now achieving higher than control-school learners. One could also anticipate that identifying and remedying deep-seated gaps in language proficiency would take longer by Grade 7. These outcomes are also depicted graphically in Figures 15-19. Again no fundamental differences were observed regarding the outcomes reported respectively by the HSRC and sJsk. These findings represent improvements ranging between 80% and 460% for learners in project schools compared to changes from 30% to 70% for learners in the control group.

**Table 19: Mean percentage achievement (age estimate for Schonell test), percentage point change, difference-in-difference and percentage change figures for Grade 7 learners by school type**

Test	School subset	Baseline % (& SD)	End Year 1 % (& SD)	Change % pts (& SD)	Diff-in-diff % pts (& SD)	Change % (& SD)
<b>Schonell Reading Age (not %)</b>	<b>Project</b>	7,2 (0,9)	***8,4 (1,4)	***1,2 (1,2)	0,7 (0,9)	16,6 (1,3)
	<b>Control</b>	*7,4 (1,1)	7,9 (1,1)	0,5 (0,9)		7,0 (0,5)
<b>Literacy 1 MCQ</b>	<b>Project</b>	35,7 (28,8)	***63,9 (24,2)	***28,2 (34,1)	16,8 (0,6)	79,1 (1,0)
	<b>Control</b>	*39,8 (30,2)	51,2 (29,9)	11,5 (39,1)		28,8 (0,4)
<b>Literacy 3 Write Paragraph</b>	<b>Project</b>	10,8 (21,1)	***47,8 (34,2)	***37,0 (38,9)	27,1 (1,3)	344,0 (1,8)
	<b>Control</b>	**15,2 (21,6)	25,1 (25,4)	10,0 (29,0)		65,7 (0,5)
<b>Literacy 4 Open-ended</b>	<b>Project</b>	4,9 (13,5)	***27,5 (23,8)	***22,6 (24,2)	17,9 (1,3)	462,7 (1,7)
	<b>Control</b>	*6,6 (12,2)	11,4 (15,9)	4,8 (16,8)		71,9 (0,4)
<b>Literacy Total</b>	<b>Project</b>	15,3 (13,2)	***42,2 (19,7)	***26,9 (19,2)	18,6 (1,4)	176,0 (2,0)
	<b>Control</b>	**18,3 (13,9)	26,7 (17,0)	8,4 (17,3)		45,8 (0,6)

\* p<0,05    \*\*p<0,01    \*\*\*p<0,001    (No corrections calculated for this report for effect of cluster sampling)

**Note:** To explore the extent of the effect of intervention status (project vis-à-vis control school), a basic value-added regression model was run. It regressed learners' November scores on their baseline scores (February) and intervention status. As explained throughout, the intention is to explore such quantitative,

technical and statistical matters further in future, preferably closer to the final evaluation and in academic outputs. Hence, no additional tables are included. Some core statistics appear in an endnote<sup>iii</sup>.

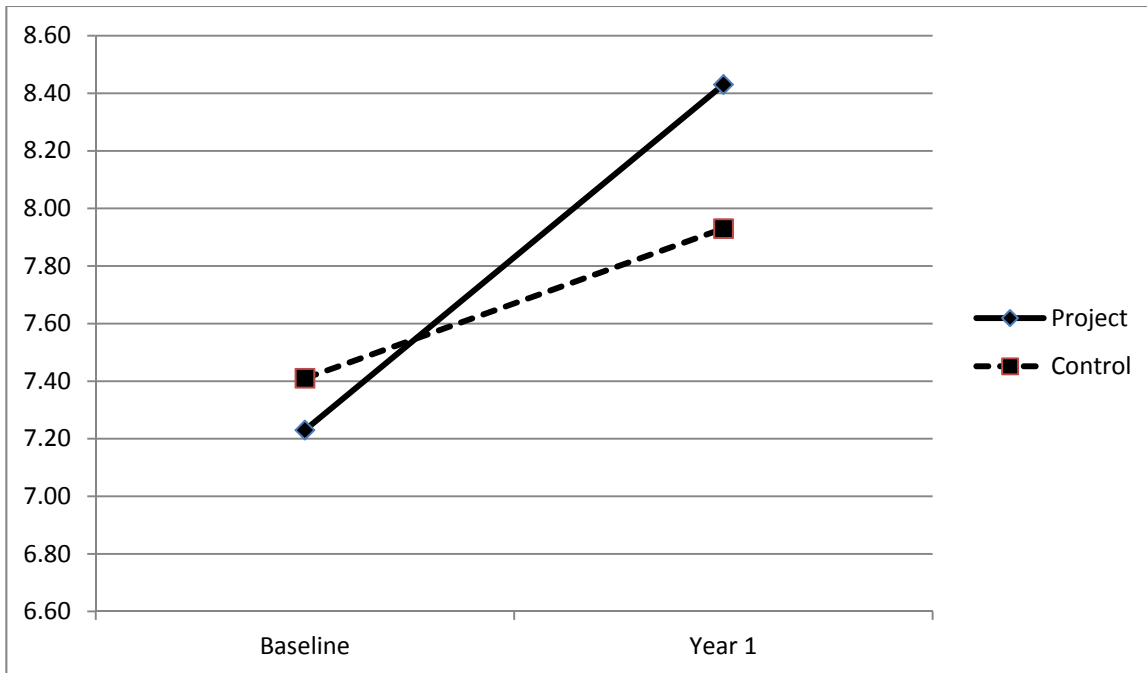


Figure 15: Grade 7 mean learner scores (age) on the Schonell Reading Age test over time by school type

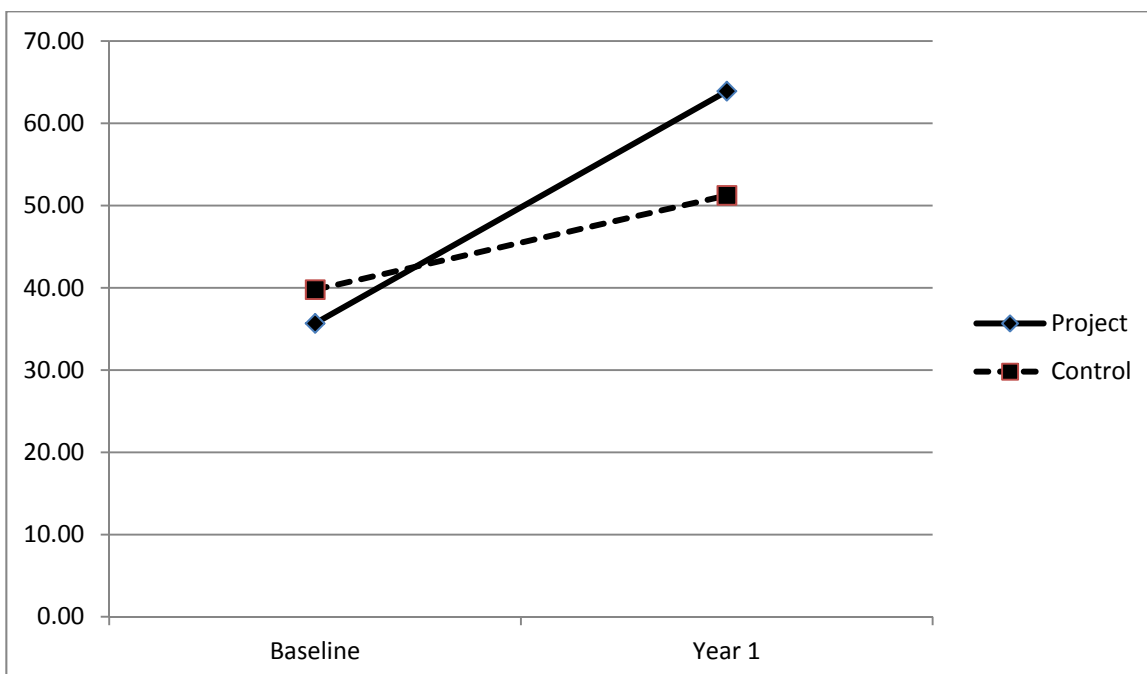
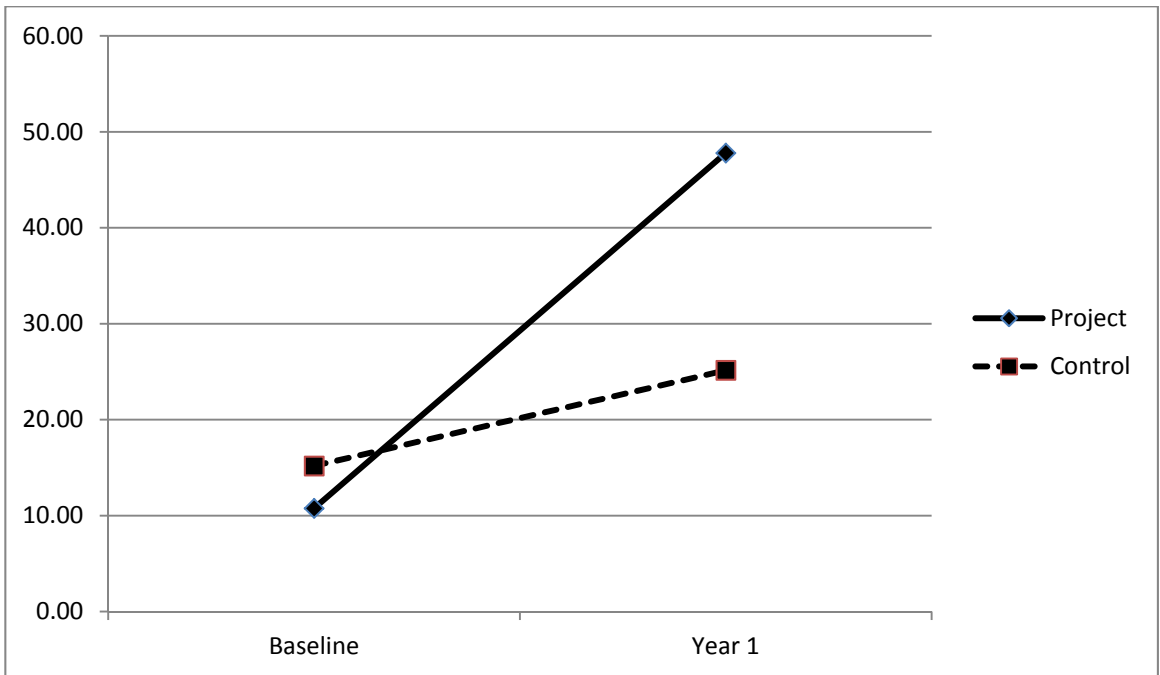
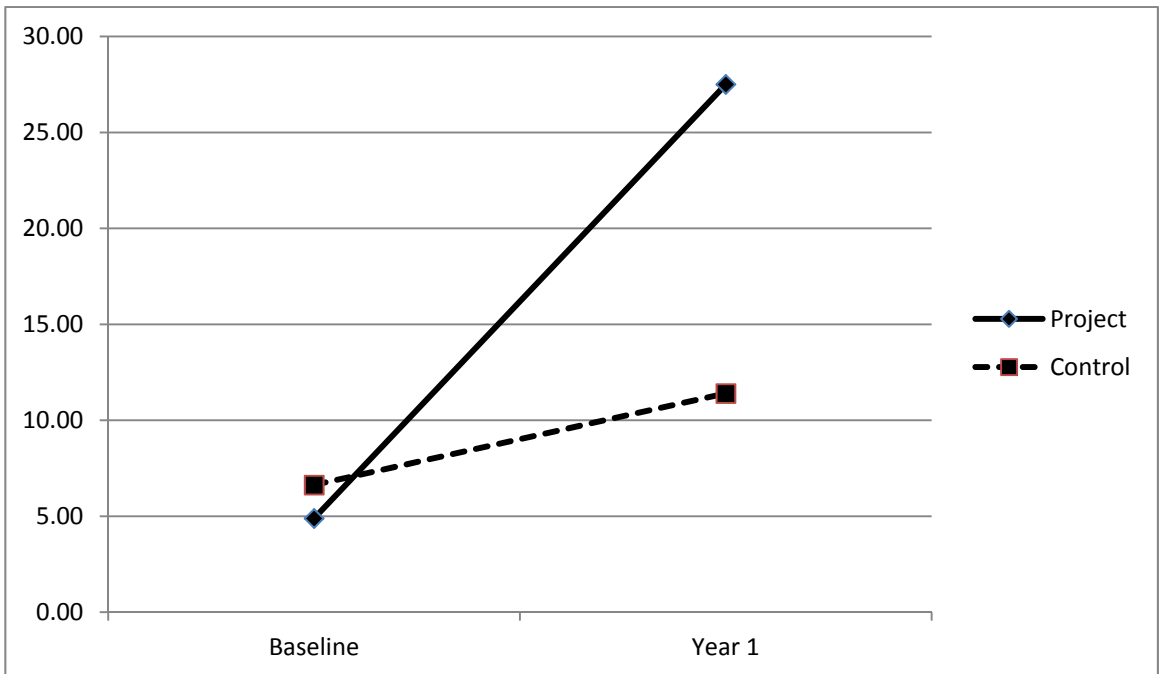


Figure 16: Grade 7 mean learner scores (%) on the Literacy 1 sub-test (MCQ) over time by school type

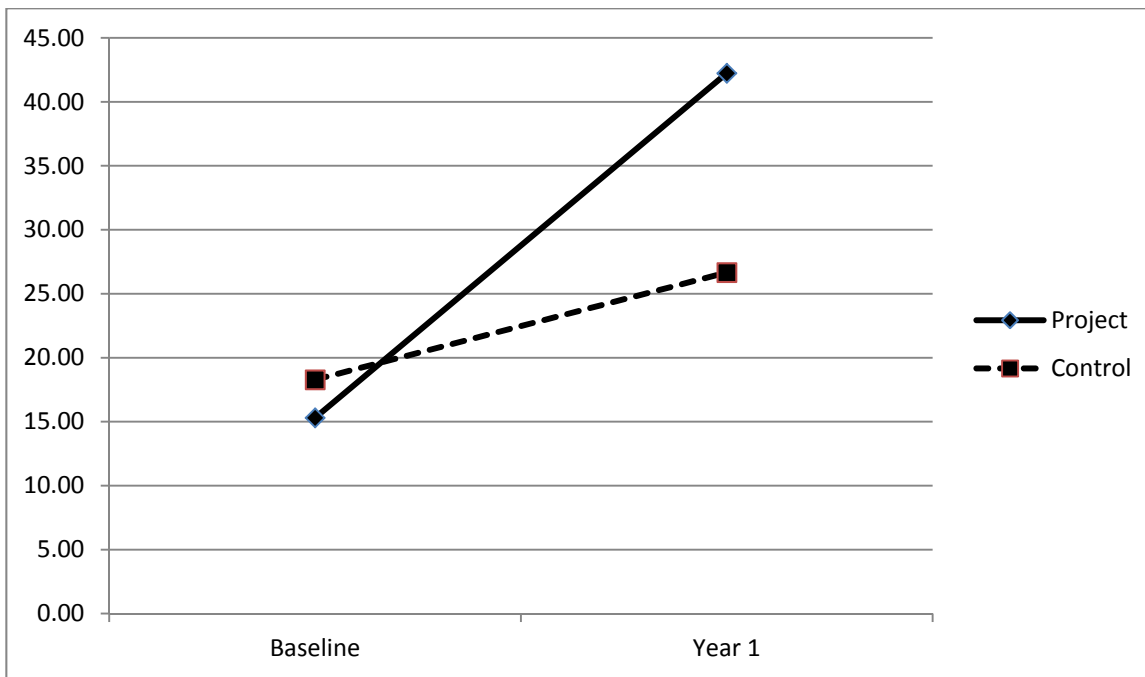


**Figure 17: Grade 7 mean learner scores (%) on the Literacy 3 sub-test (Paragraph Writing) over time by school type**



**Figure 18: Grade 7 mean learner scores (%) on the Literacy 4 sub-test (Open-ended items) over time by school type**





**Figure 19: Grade 7 mean learner scores (%) on the total Literacy test over time by school type**

As for the other two grades, achievement increases and changes are also compared at this point in terms of the percentage of learners at baseline and after Year 1, for project and control schools, who “mastered 50% of their work”. Table 20 reflects the situation. From there it can be seen that the initial more favourable position among learners in control schools in February was equalled by learners from project schools by June and significantly surpassed by them by November across all tests. Only one statistically significant set of scores existed between learners at the outset. Control-school learners were outscoring project-school learners in paragraph writing, but only at the 95% probability level. No significant differences remained between learners from project and control schools by June. By November, after one year of intervention, significantly more project-school learners were achieving at least at the 50%-mark on all the tests.

**Table 20: Percentage of learners in project and control schools achieving a reading age of 11 or higher or at least 50% in the three Grade 7 literacy sub-tests and its total score during testing in February, June and November 2013**

Test (n)	School subset	Baseline	June	November
Schonell Reading Age of 11 yrs (n=874)	Project	0,6	1,4	**6,3
	Control	1,1	2,4	2,0
Literacy 1 MCQ (n=936)	Project	30,7	43,2	***73,6
	Control	36,6	44,3	52,3
Literacy 3 Paragraph writing (n=924)	Project	13,2	12,8	***59,9
	Control	*18,2	16,9	32,0
Literacy 4 Open-ended items (n=930)	Project	1,2	1,6	***20,8
	Control	1,7	1,8	4,6
Literacy Total (n=936)	Project	1,9	3,2	***35,3
	Control	3,4	3,0	10,8

\* p<0,05    \*\*p<0,01    \*\*\*p<0,001    (Correcting p-values after cluster sampling has not been done)

Note: p-values based on Phi and Cramer’s statistics

It should be acknowledged that it was a daunting task to get even project-school learners to read appropriately to their age in the course of a year, given the performance lags that emerged among Grade 7 learners over many years. The effect of regular public schooling in the control schools was an increase of only half a year in learners’ reading age in the course of a year. They therefore fell further behind, now to over three years. As a result of the intervention programme, the reading age of project-school learners

increased by 1,2 years, which is at least more than the required minimum of one year. However, they were still more than two years behind the required level. Put differently, after a full seven or eight years of schooling, only between 2% and 10% of the learners are at the required level. The pace of catching up should clearly be higher. It can be noted, though, that the Grade 7 group was in the programme for only one year. The Grade 4 cohort of 2013 made much stronger inroads into their reading backlogs, and the fact that they continue as a cohort bodes well for such a group to catch up by Grade 7. In addition, the intervention programme seems to have succeeded in starting to address some crucial elements of learner comprehension and writing ability. Tables 19 and 20 provide the necessary details underpinning these final comments.

### **3.3 Comparing learner achievement between monitored and not-monitored test administrations**

As explained upfront, HSRC staff would monitor (“quality assure”) a higher than usual portion of assessments at schools. That would serve as mechanism to offset any unforeseen or undesired effects arising from the decision that sJsk, and not HSRC, would conduct the assessments. Assessment, scoring and data capture by sJsk were required to ensure that sJsk had immediate and regular access to the results. Learner assessment took place from 11 to 14 February 2013 (i.e., the baseline assessment), from 4 to 6 June 2013 (i.e., mid-term or Semester 1 assessment), and from 28 October to 6 November 2013 (end-of-Year 1 assessment). February-to-June 2013 comparisons were made in the interim or provisional evaluation report covering the intervention impact over the first semester of the programme. However, the availability of full-year data now warrants focusing on February-to-November 2013 learner-achievement comparisons. Only brief mention is made to observations from June 2013.

Having monitored some assessment events, one can link the status of a learner or class group, that is in terms of having been monitored or not, to their learner achievement marks. One then usually investigates whether or not there are different rates of improvement in achievement for learners on the basis of their monitoring status. This can be extended to focus also on the assessment marks as such at the baseline or after the period’s intervention implementation, again in relation to monitoring status. The rationale for doing such analyses is that test administrators may, if not deliberately, even sub-consciously “rig” marking activities and scoring standards to ensure that learners in control schools would not seem to gain much benefit, while those at project schools would. However, in practice this is a very difficult thing to get right. Not only must baseline scores for control-school learners be inflated at the start but suppressed at the end, and the inverse done for the scores of project-school learners, but this also has to be maintained realistically across grade groups and over moments of assessment for all cohorts over three years. Subconscious, rather than deliberate, effects are thus more likely. Effects can range from being erratic or inconsistent to being systematic, and from negligible to substantive. The purpose of the discussion in this section is to determine and comment on any perceived effects. In order not to distract from the main impact evaluation findings, it was decided to keep the section as brief and as meaningful as possible. This was pursued in two ways. First, the immediate focus was placed on a qualitative discussion of the monitoring observations. Second, the quantitative figures were retained as annexures, with only the briefest possible comments in the text of the three sub-sections that follow.

Before reporting the findings from monitoring, some background is given on how monitoring took place. The extent of monitoring is reported on first. Over the course of three to six days each time in February, June and October 2013, some schools on sJsk’s normal assessment schedule were visited by a team of three HSRC staff. In the case of control schools during November 2013 only two monitors paid such visits. This resulted from programme changes and staff availability, mainly as a result of service-delivery protest action in the Malamulele area making assessment activities and monitoring visits unsafe. Two monitors each time followed as many testing events as possible at a school. When available, which was in most cases, the third member organised and observed overall logistics and also visited other schools to fill gaps in contextual information. The observers recorded which grades’ and classes’ test administration was observed, and noted any other qualitative information on the prescribed monitoring sheet. As far as possible, all test contents were covered in the process for a grade / class groups. As a result, all of a learner’s assessment contents (sub-test activities) were considered to have been monitored each time. At

the baseline, after the first semester and after a year, when calculated on the basis of the number of learners in sessions that were monitored, monitoring coverage was as reported in Table 21. The proportion of learners who were in both the February and October/November monitoring sessions were 24%, 19% and 18% respectively for Grades 1, 4 and 7, and only came from project schools. This was not planned, that is, these schools or learners were not sampled as such, but it was a mere fortuitous outcome based on logistical and practical possibilities concerning the visit programme.

**Table 21: Monitoring coverage at the baseline and after the first semester and first year**

School type	Gr.	Baseline: February 2013		Semester 1: June 2013		Year 1: Oct/Nov 2013	
		% monitored		% monitored		% monitored	
		Overall	Per school type	Overall	Per school type	Overall	Per school type
Project Control	1	(35%)	37,9	(12%)	21,5	(60%)	64,7
			28,4		00,0		47,0
Project Control	4	(32%)	32,2	(11%)	18,1	(60%)	64,6
			30,5		00,0		49,0
Project Control	7	(30%)	30,4	(34%)	54,3	(74%)	85,0
			28,3		00,0		49,5

HSRC staff were trained beforehand in completing the monitoring forms at the assessment sites and submitting them after the visits. These forms were designed to establish how closely sJsK facilitators followed the administration manual developed by Ms Pat Murray of sJsK . It was titled: *“How to Conduct Learner Assessments”*. It set the standard according to which the learner achievement assessments should take place and served as training manual for the test administrators beforehand. The monitoring forms also allowed for recording consistency of scoring the various individual oral and written group tests and recording of results. Scoring that happened at the individual learner level as part of oral assessment at schools was not monitored on the same basis as marking written scripts at a central operations venue afterwards. For the component of scoring that took place after school at a central operations venue, learners may not always have been linked directly with the records of such bulk marking events that had also been monitored from time to time.

Monitoring forms covered the following information: identification details pertaining to date, grade, class, school, monitor and test administrator; accuracy of following the instructions script (and assessing the impact of any deviations on marks); setting learners at ease; presenting the correct assessment items to every learner; correctly recording school, learner and test details; refraining from coaching or helping learners during assessment; accurately marking and scoring the individual items of the assessment; refraining from correcting learners’ mistakes or giving them discouraging criticism; showing patience and communicating understanding during the assessment; noting particular test-related observations; keeping test materials secure during, before and after the assessment; confidence and preparedness of test administrators; the extent to which learners coped with the procedures; factors or circumstances that may have influenced the achievement of learners during the assessment; conduciveness of conditions under which testing occurred (venue, site, room, lighting, temperature, noise, etc.), i.e., any serious distractions; and overall quality of the testing session (five-point scale rating).

In-depth analysis of the submitted monitoring sheets, especially up to and after the first occasion during June 2013, was done by the doctoral intern<sup>17</sup> attached to this project. It revealed that by and large test administration procedures were followed very tightly. Excellent standards were considered to have been maintained in most cases. Else, standards were considered to be very good or at least good, but never problematic. Scoring was considered to be excellent. It was once noted that some test administrators had not yet reached consistency in their scoring, which resulted in a few incorrect scores in the early stages. During October/November 2013 it was clear that consistency of test administration and item scoring reached very high levels. The assessment manual script was followed very closely and accurately in virtually

<sup>17</sup> Ms Shawn Rogers’ processing of the contents of these forms is hereby acknowledged.

all cases. This could be ascribed to the implementation team's familiarity with the learner assessment procedures and assessment contents. If anything, mechanistic or tentative administration of the assessment may have crept in owing to the routine nature of some activities, but that was only observed once. The programme leaders constantly and regularly encouraged and motivated the assessment administrators.

The few concerns and challenges, largely in a small minority of cases, still observed by June 2013, were not considered to have influenced the quality of test administration in a systematic way. They were summarised at the time, as reported below, to enable sJsK to maintain their high assessment standards. This sJsK did through regular refresher sessions with administrators before subsequent assessment cycles, a practice that was indeed continued into the second semester. The evaluation team also had to ignore some initial comments made by monitors after the baseline and first-semester visits pertaining to how the Letter Sound sub-test was administered and scored, as such comments would contradict deliberate linguistic or other approaches followed in the intervention programme, which monitors may not have understood at the time. (Administrators had to ask for the sounds represented by letters, and not their alphabet names.)

The structure already pursued above will be retained during reporting in the few paragraphs below. This refers to first reporting the situation during initial baseline and first-semester monitoring before commenting on any changes or making additional observations on the basis of end-of-year monitoring during October/November 2013.

Initially (at the baseline and the end of the first semester), it was observed that some test instructions were not well communicated to learners. In an isolated case, it was noted that the pace of one administrator in administering the Non-Word Recognition test may have been too fast, not letting learners attempt it properly. Also in a single case, regarding the written Comprehension sub-test, the story was read to learners, which was a deviation from the prescribed procedure where learners had to read the story themselves. In another case, administrators did not emphasise that learners had to read through the passage carefully first, and some learners started answering questions without reading the story. Also, one monitor wondered about the time allocation of an hour for the written Comprehension test being too little for some learners. In one more case, the only instructions given were for learners to fill in their school name, grades, date and own name, without making them feel at ease. During administration of the Letter Sound test, some learners were found not to fully understand that they had to sound the letters. They did not follow fully what "sound out the letter" meant, even though this was explained numerous times. (Note: This may be a case of confusing either or both the intervention objective and /or test task in terms of alignment with either merely naming the letters of the alphabet or focusing on their phonetic sound characteristics.)

By November, occurrences of the preceding outcomes had largely ceased. Test administration was mostly considered to be very good, or at least good, and even excellent in a few cases. Learners were considered to be at ease, and dealing with assessment as a normal activity. Very few individual learners did, however, struggle to read, appear very timid or lack confidence. This seemed to be linked to individual proficiency and temperament, and not assessors' assessment approach. Some Grade 1 learners in the process still "froze" or locked up totally. This, however, would not be strange at such a young age, and the test administrators, under sJsK's guidance, dealt with it well by not extending pressure unduly, but by providing learners early opportunity to opt out of the situation. One Grade 1 learner was seen to start crying in the process, but this could have been related to other discomforts on the day. Learners were largely considered to be at ease, and often described as happy, unafraid and brave. In two isolated cases invigilators read or repeated the instructions to the written tests after learners started reading or answering the test. This may have distracted some learners. Other very isolated and slight deviations in single cases are briefly noted next. A single invigilator was allowed once to administer a written test session when staff allocation to competing activities and a sufficiently small number of learners in class allowed. Here and there administrators may momentarily have lost concentration and marked isolated Schonell word pronunciations as correct. A new test administrator initially confused Word Recognition sub-test and Schonell administration procedures. This was detected soon. Scoring of the few already completed test

answer sheets was corrected immediately, and the administrator was coached back to the right procedure. None of the foregoing is considered to have jeopardised the reliability of scores and the high quality of assessment otherwise.

Oral assessments are done outside in the open, with five to six or more administrators to a table during assessment. Constant talking by administrators and learners occurs there. Some monitors at the baseline and June rounds questioned the appropriateness of this approach given the possible disruption of the learners. However, they also noted that learners did not copy the answers of or listen to other learners. This situation remained unchanged during November, and cases were very few and far between where learners seemed to be distracted by neighbouring events or responses. A situation was spotted once where two learners were not separated by one of their administrators, because of the inconsistent implementation of the arrangement of letting learners all stand on the same side of their administrator. This was corrected quickly. It would be considered prudent for test administrators to continue respecting learners and ensuring they are at ease by minimising occurrences, granting that such were very isolated, of potential distraction to learners. This would include reducing laughter and talking between administrators when learners are around, not ordering or pointing learners about in too impolite a manner, and not exposing learners too freely to ticking off their verbal answers as correct or incorrect in their full view. In addition, test administrators should refrain from giving subtle signals that a learner's answer may not be correct, prompting them thus to make a corrective attempt, or, once marking a response as incorrect, actually telling them what the correct response would be. The latter practice could assist them in gaining a higher score next time. However, such conduct did not occur often.

Furthermore, the home language was (and may be) needed for explaining instructions. The fact that instructions were given in English seemed to make some learners hesitant at first. Some of the learners did not understand how to complete the multiple-choice section before they started. The abovementioned situation, and necessity, that prevailed at the baseline and in June continued into November only in the case of the Grade 1 oral testing. These Grade 1 learners were often set at ease and organised using their home language. The appropriate assessment contents and procedures, though, were not compromised.

Before the November monitoring, limited school infrastructure and conditions were considered to hamper optimal assessment in a few cases. These would include the number of desks for the size of the class, resulting in learners being seated three at a desk, or otherwise sitting too tightly. Here and there the venue was poorly lit or quite dark (including electricity not being connected at all). Testing was once affected by a neighbouring school making noise with drums. Testing once coincided with a school break, so that there was much noise and curious learners seemed to distract those being assessed. Possible distractions were similar in nature, but much less pronounced, during November monitoring. Twice very overcrowded classrooms were the case during written testing. It was addressed well in the case of a particularly large group of learners by sending four administrators to conduct the session. Humid hot weather was noted in four instances to make conditions difficult. Written assessment was fortunately completed in most cases by 12:00. A gloomy and overcast morning, linked with the absence of electricity, made classrooms a bit dark. However, once everyone's eyes adjusted to the available light, conditions were not considered to hamper assessment. In one instance where other learners made excessive noise outside, it was possible for the administrator to get them to move elsewhere.

As alluded to above, learners were aware of the use of the "x" by facilitators to mark incorrect answers. This seemed to distract some of them when points were assigned. This situation continued for the duration of the year. Would there be a way of using different symbols, or for learners not to see the marking?

Administrators from the beginning did not conspicuously have or use memorandums for scoring tests. They appeared to have memorised the correct answers. This could become more of a problem in marking the Comprehension sub-test. Markers were observed to work in groups and discuss the "correctness" of the answers offered. This process was monitored by their supervisors. Monitors wondered about this arrangement being optimal and consistent. (The service provider (sJsk) clarified their approach in this regard after a first version of the previous five lines of this paragraph had been written in June. Being rather

short instruments, which mostly used multiple-choice items, markers indeed quickly memorised the correct answers. However, that was based on the existence of formal memorandums and ample training until acceptable proficiency levels had been reached. Additional moderation was always done in particular in relation to the open-ended responses, because new learner responses are continuously produced. Markers needed and got ongoing updates to identify and agree on such acceptable answers.) These on-site practices related to marking the written tests afterwards at a central venue were again monitored at two occasions during November, and no deviations or inconsistencies were observed.

In conclusion, on also analysing monitors' feedback after November 2013, test administrators can be said to have dealt with their assessment task very well and consistently. They also overcame conditions of under-resourcing at schools in the best possible way. No significant factors were observed that could jeopardise the reliability and validity of the evaluation information or intervention implementation in any significant way.

The initial overview after June 2013 of the association between sessions having been monitored or not and learner achievement was based on a rather short time lapse since the commencement of interventions, and thus also on a low intervention volume. As a result, the data could not be expected to be stable yet. However, after a full year the current situation is very acceptable. It has to be kept in mind, still, that linking monitoring (at a site and for a grade, class and test instrument on the whole) to every individual learner's item, sub-test or overall test scores is not possible given the pace and extent of the assessments.

Overall, on the basis of being monitored at either the baseline, or after one year, or on neither or both occasions, the outcomes in terms of learner achievement scores and gains over time are reflected in Annexures 4 to 6. Although it can again be mentioned that it is not advisable to overly rely on these statistical indications, they are provided in the Annexures to give an indication of the spread of scoring or assessment trends. Had corrections for the clustered design feature of the study been done, a majority of the "significant" indications would not have been made. To assist in identifying such trends, those that would possibly constitute biased scoring are indicated by **shading** the asterisks associated with the relevant probability levels (e.g., as **\*\***), while those totally counterintuitive to a possible influence of scoring to make effects look stronger are underscored (e.g., as \*\*). This enables giving a brief summary of key observations by grade in the sections that follow. Therefore, what follows below is a selection and summary of salient observations, given the vast volume of resulting calculations and figures. Not nearly every comparison and statistic can be reported fully. However, analyses were conducted rather fully and systematically. As above, the findings from the latest full-year comparisons are preceded by brief indications about the situation after the first-semester comparisons by June 2013, without reporting those statistics. Outcomes in terms of specific learner score levels in February and November 2013 are not discussed unless there were very noteworthy patterns. Learner gain scores between these two points were the focus, rather.

### 3.3.1 At Grade 1 level

Only two possibly meaningful (significant) findings were yielded in relation to learner achievement gain scores over time as evaluated by June between monitored and not-monitored schools for project-school and control-school learners. They both pertained to the Oral Pictorial test, and were both counter-intuitive to what would have been attempted had there been efforts at manipulating effect patterns. For control-school learners on the basis of February monitoring, the gains were lower at monitored sites (\* $p < 0,05$ ). For project-school learners on the basis of June monitoring, the gains were higher at monitored sites (\*\* $p < 0,01$ ). As expected, the latter finding was also yielded from analyses using partly overlapping monitoring records accounting for being monitored on both, either or none of the occasions.

The main observations that can be made from Annexure 4, on the basis of the full-year November data, are listed next. They are provided in point form as the details appear in the annexure, would be too many to repeat in the text, and would give a false sense of finality given the preferred qualitative focus.

*For achievement scores and gains among learners in the project group*

- Score levels and gain scores are inconsistent and contradictory, that is, not supporting any systematic assessment or scoring biases, between and across sub-tests (especially for November scores), across assessment dates (February and November), and on the basis of being monitored in the different semesters.

*For achievement scores and gains among learners in the control group*

- By far most of the observed differences are counterintuitive; they indicate opposite trends to what would constitute assessment or scoring bias attempting to show greater intervention effects.
- As with the project-school assessment data, but more so, score levels and gain scores are inconsistent and contradictory across sub-tests, assessment dates and monitoring points.
- The observed baseline (February) achievement scores and the score gains from that basis often portray opposite “effects” in terms of evidence towards any assumed assessment or scoring bias.

### 3.3.2 At Grade 4 level

In relation to learner achievement gain scores over time as compared between monitored and not-monitored schools across project and control learners by June, only one significant relation (out of 24 possible total relations) was of logical or practical meaning. It pertained to February monitoring linked with overall gains at project schools for the Literacy 4 sub-test. In the remaining project-school cases, children from sites monitored during February actually gained more. Three significant findings were made from control-school comparisons on the basis of February monitoring, one for project-school sites on the basis of June monitoring, and two for project school sites on the basis of comparing outcomes for being monitored on both or either of the occasions, or none. These were all counter-intuitive.

Main observations on the basis of Annexures 5a and 5b for the full-year November monitoring data follow.

*For achievement scores and gains among learners in the project group*

- Observed achievement score patterns are quite inconsistent (i.e., random?) between monitored and not-monitored learners across assessment points (February and November). A further inconsistency concerns the sub-tests. For the EGRA sub-tests the November scores particularly lack signs of assessment or scoring bias, while for the Literacy sub-tests it is the February scores.
- Data trends are contradictory across monitoring point data (February, November, either or both).
- Data trends are inconsistent or contradictory across sub-tests. This and the previous point apply in particular to gain scores over time in relation to the EGRA sub-tests and Schonell Reading Age assessment.

*For achievement scores and gains among learners in the control group*

- A large majority of differences on the basis of monitoring status are counterintuitive, especially for gain scores over time and November scores, and in particular for the Literacy sub-test scores.
- Observed achievement score patterns are rather inconsistent across monitoring points.
- Score patterns are contradictory and inconsistent across assessment points (February and November).
- Data trends are inconsistent or contradictory across sub-tests.

### 3.3.3 At Grade 7 level

At the time of the June evaluations, comparing learner achievement score gains between learners from monitored and not-monitored assessment sites on the basis of monitoring records from June yielded no significant findings. Only one significant finding was made when doing the same on the basis of February monitoring records. In that case the Literacy 4 subtest (Comprehension) showed a very small gain score for not-monitored learners compared to a decrease in achievement over time among monitored learners. This is actually counter-intuitive and not of practical or logical significance. The Literacy 4 sub-test and Total Literacy scores of learners from sites which were monitored on both occasions showed a significant decrease over time, instead of remaining constant as for those monitored once, or improving slightly as for those not monitored ( $p < 0,010$  in both cases). Project-school learners were at stake on both occasions.

Main observations on the basis of Annexures 5a and 5b for the full-year November monitoring data follow.

#### *For achievement scores and gains among learners in the project group*

- Observed achievement score patterns are quite inconsistent between monitored and not-monitored learners across assessment points.
- Data trends are inconsistent or contradictory across sub-tests.
- Data trends are contradictory across monitoring point data (February, November, either or both).

#### *For achievement scores and gains among learners in the control group*

- A very large majority of differences on the basis of monitoring status are counterintuitive, especially for gain scores over time and November scores.
- Score patterns are quite contradictory and inconsistent across assessment points.
- Data patterns across sub-tests are inconsistent or contradictory, with a view to possible assessment or scoring bias as always above, especially for assessment scores from February.

The final conclusion is that patterns and effects across grades, (sub-)tests, assessment points and monitoring points are too erratic, inconsistent and contradictory, not to mention the majority of them identified as counterintuitive, to constitute detection of any deliberate attempts at assessment or scoring that would make the intervention look more positive than it was evaluated to be. The observed patterns, quantitatively and statistically speaking, are likely to be random. As indicated at the outset, it would take a conscious, concerted and consistent effort to ensure that assessment scores are realistically distributed (manipulated) in such a manner that low initial scores among unmonitored project learners would be turned into high scores over time, relatively speaking, or the inverse among unmonitored control-school learners starting off at relatively high score levels but not gaining much over time. The evidence at hand confirms that assessment and scoring have run without bias as being monitored has not changed regular assessment or scoring patterns substantively, consistently or in a purposive manner.

In further support of the conclusion above, two additional explanations may support prudence in ascribing any observed trends too hastily to assessment or scoring bias aimed at achieving a favourable evaluation of effect. The first is the concise nature of the scales. The low number of items to many sub-tests restricts score ranges and enhances the effect of any individual item score on sub-test totals. The result would be larger fluctuations than usual, which one could easily wish to attribute to deliberate scoring bias, although possibly still largely random. The absence of large and consistent effects in the current data provides no ground for a conclusion other than the one made above. The second issue is about the dual effect of being



monitored. It may, on the one hand, remove deliberate or unconscious bias in scoring or assessment. On the other hand, it may also merely increase concentration and scoring quality in general. The effect of the latter could be oppositional in many cases, and thus increase inconsistency further. The opposing scoring quality effects could entail that assessors and scorers miss either wrong or correct answers when not being monitored. Being monitored may return optimal motivation and improve assessment standards. The two points made, therefore, further caution against considering the presence of scoring bias especially when other known factors may increase inconsistent trends.

Note: In addition to initial exploration through a basic value-added regression model of the extent of the effect of intervention status (project vis-à-vis control school) on programme outcomes, robustness checks were also run by adding monitoring status and its interaction with intervention status. The resulting changes to the volume of variance explained by the initial value-added model and to the standardised coefficients for intervention status are also noted in the three endnotes referenced earlier. The intention is to explore such technical and statistical matters further in future, preferably towards the final evaluation and in academic outputs. Without producing detailed tables, core statistics appear in the endnotes.

### **3.4 Uptake factors and systemic issues influencing rolling out and scaling up the intervention**

Notes were made all along on thinking related to process and dynamics during starting up and rolling out the intervention implementation. These are now structured to allow a provisional evaluation of matters that need to be heeded as the programme expands into another year, a new cohort and more schools. These matters will be of special relevance in future on considering larger provincial and national scaling up.

It became clear very early that much pre-consultation and conversation were absolutely necessary, and in fact had taken place, to ensure that the intervention implementation, and its accompanying evaluation component, started off on a solid footing. Before providing some detail, it is concluded that the start-up phase was a very positive experience. Some main elements are briefly highlighted next.

The importance of the existing networks of sJsK cannot be underestimated, as these largely led to the identification of the Mhinga area as the eventual location for the first cohort of the study.

Chiefess Lorraine Mhinga *de facto* started serving as the “patron” of the programme early on in 2012. On the basis of her already existing social responsibility programmes in the areas of health, HIV/AIDS intervention, early childhood development and establishing a community library, to name a few, she undertook to support the sJsK learner intervention by all manner and means. That did not only entail giving access to the tribal office and other facilities from time to time, but also sponsoring early stakeholder meetings, entering into conversation with provincial education officials, and the like. A lot of canvassing for the intervention and study took place in this way.

Through the mediation just referred to, an official request to the provincial office of the Department of Basic Education (DBE) to approve the study was quickly met with the necessary formal approvals by and support of both the provincial and the circuit offices of DBE.

The circuit manager of Malamulele North-East, as a result, also gave the project his unflinching support. This became an important factor gaining and retaining access to the relevant schools afterwards.

In addition to many other direct communications, two main formal meetings put the programme on course. The first was the large stakeholder information meeting in Mhinga on 15 October 2012 at which sJsK, DBE, USAID and HSRC all provided background to the intended programme. In attendance were representatives from the broader schooling system, including circuit staff, principals, management team members and teachers from the project and control schools, representatives from teacher labour unions, members from the Mhinga tribal authority, community members, and a few others.

The second meeting was a high-level official launch ceremony in Pretoria on 8 November 2012. It covered not only the sJsk intervention, but also two others funded as part of SCIP. At this occasion, the Deputy Minister of DBE (Mr Enver Surty), the US Ambassador (Mr Donald Gips), other representatives from USAID's mission in Pretoria, representatives from ELMA Foundation (among them Ms Bernadette Moffat as Executive Director from Elma Foundation Africa), JP Morgan Chase (Mr John Coulter, Senior Country Officer for Sub-Saharan Africa), sJsk, HSRC, etc. were all in attendance to celebrate the announcement of the various programmes and learn more about them.

Then a period of recruitment of coordinators and facilitators by sJsk followed, still during the second semester of 2012. From over 200 community members interested in being trained for their positions, short-listing and face-to-face interviews followed and led to the selection of the final 25 facilitators.

sJsk then embarked on training them for their intervention and assessment roles. This phase ended only into 2013 once sJsk was satisfied that the required levels of proficiency had been reached. If anything, the extent of this task was much larger than anticipated, even though its importance had never been underestimated. It in effect took much more effort and longer than envisaged. Even after the Easter holidays, some retraining was required before the intervention implementation reached full efficiency.

The team of sJsk in the meantime selected, developed and prepared all the intervention and assessment materials and produced the required volumes thereof.

Other logistical preparations and infrastructure acquisition were undertaken in close consultation with the funders and in interaction with the community and its many representatives. These included the purchase of a vehicle, trailer, tables/chairs, crates, etc., and also trying out and settling on the best Johannesburg and local site accommodation and transport arrangements for the team as well as the facilitators from the community.

It also entailed opening up and equipping office space in Johannesburg, complete with all the required financial, administrative and accounting staff, systems and procedures.

We cannot neglect mentioning the sterling support afforded both the implementation and evaluation team by a volunteer at the Tribal Headquarters of Mhinga. Once the intervention process was ready to start, he provided invaluable protocol services, as it were, besides bringing us to all the sites and showing us around the area on the basis of his deep familiarity with its geography, history and sociology.

All that was left was for sJsk to finalise their intervention plans and schedules on the ground and to begin the task.

Some of the critical lessons learnt (or re-confirmed) include the following:

- Buy-in took long and required dedicated effort and attention to relationships.
- The time that role out / ramping-up was going to require (infrastructure, equipment, etc.) was underestimated initially.
- Recruiting and training the relevant facilitators also took longer than provided for.
- School functionality made many aspects of contextual base-lining difficult; one critical factor being up-to-date and available learner records and registers.
- Parental factors (such as their ability to read and write, and follow questionnaire completion instructions, the distance they live from schools, their limited response to school invitations and low general involvement levels) complicated and delayed home-background base-lining and reduced the return rate.
- Maintaining and refreshing initially established relationships and agreements through the course of the year deserved and got ongoing attention and vigilance. This countered natural tendencies inherent to projects such as this for memory to fade. Also, any new individuals coming on board through the year, perhaps because of staff changes, which the project was largely spared, had to be initiated into the many activities and their underpinning assumptions, context and knowledge.

- Flexibility in implementation always had to be foregrounded, and was. The intervention and evaluation team, sometimes at short notice, had to make contingency plans and adjust schedules and staff movements to accommodate for external factors beyond their control. Two examples will suffice. The almost tropical nature and extent of the rains received in summer, coupled with features of the terrain and access to it, such as gravel roads, sandy areas, river crossings, mountain roads, etc., make certain routes unpassable at times. Special vehicles, equipment and skills are then necessary. Service delivery protests and violence, coupled with underlying tensions of politico-ethnic nature (closely related in cases to living in Tshivenda- or Xitsonga-dominated regions), and complicated by the ease with which single critical access routes or intersections could be and were in fact used for controlling movement through the area, kept the intervention and evaluation teams out of the area during a spell of almost two weeks during October/November 2013.

A SCIP “Learning Forum” workshop was organised by REOS on 25 November 2013. SCIP partners, including implementing partners, the DBE and other relevant actors shared information and engaged in discussion about the challenges around improving primary-grade reading results in the country. Participants were required to focus on success, what was not working, areas of improvement, systemic challenges and effective communication and dialogue. In line with what is reported on elsewhere in this report, although discussions at the SCIP event was not intervention specific all the time, sJsK also shared their experiences from the field, and learnt from those gained by the other service providers. This event is judged to have played a positive role in re-energising intervention teams at the right moment before embarking on the expansion of their first year’s work into the second year on the basis of critical reflection.

Information collected at the circuit level points to a few additional concerns that may come into play when expanding the programme to more and more schools, especially when involving additional circuits and districts. One of them is that a number of language subject advisor posts are vacant. These appear to be mostly at the senior primary school level, that is, from Grades 4 to 7. They involve both English as First Additional Language and Xitsonga as Home Language. Latest information suggests that no such posts are actually filled at present. The point is that for schools to receive optimal benefit from system monitoring and support, such posts have to be filled by skilled staff. Only then can one begin to ensure that the impact of expanded (or expanding) interventions is sustained. Partly related to this point is that circuit offices may struggle on occasion to dispense of their functions in relation to schools optimally. This could partly result from having too many, that is almost half, of their administrative and support staff being considered as inappropriately or insufficiently qualified. Finally, improving some of the mentioned situations may allow circuits to increase the quarterly frequency with which they currently report being able to visit schools and their language teachers.

All the lessons documented above are relevant when the second cohort of learners and new project and control schools, and including another regional language, that is, Tshivenda, are being introduced in 2014 to the intervention implementation and evaluation.

## 4. Discussion

### 4.1 Conclusions

*Are the conditions at project and control schools equivalent?* Our considered opinion is affirmative given the thorough analysis done and reported on above. These analyses also had as purpose describing the context in a much deeper sense in terms of its demographics and other sociological dimensions in order to ensure that the intervention implementation could be as relevant as possible. This would not only be an important source of information when expanding the project into 2014, but also serve similar new programmes well in future. It also served as motivation for retaining Section 3.1 as a longer version than seems required.

*Has the sJsk programme yielded a promising impact thus far?* This is indeed the case, given the many significant improvements across sub-test scores when compared across learners from project and control schools. The evaluation followed a deliberately strict comparison of achievement results over time. Records were retained for analysis during this round only when data were available from both the baseline (February) and the full-year (November) assessments for a given learner. This decision may have resulted in a selection bias through differential attrition. However, comparisons conducted to establish the presence of such bias confirmed that it had not happened. Attempts were also made already to obtain at this stage external indications of initial transfer of basic conceptual repair among learners to their broader school work. However, the availability of ANA scores for both 2012 and 2013 only for the Grade 4 cohort of 2013, and the quality of the data that could be obtained under short lead times, did not provide meaningful coverage or results. This matter will be kept on the radar.

*Were the assessments administered and scored consistently?* This firmly appears to be the case, given the results from detailed analysis investigating the equivalence of assessment marks across monitored and not-monitored assessment. In a majority of cases the differences between monitored and not-monitored sites were insignificant and even random or erratic (unsystematic). Where significant differences were observed, these were very often inconsistent and often counterintuitive. The latter means that a high portion of such significant findings reveals effects that are contrary to assumed (tested for) subconscious or deliberate efforts at influencing assessment scores in such a way that they would appear to show that interventions had a greater effect than it actually had. Analyses and conclusions took into account that baseline and November (also June) scoring for learners from project and control schools had to be manipulated, as it were, in four (if not six) different ways to consistently influence the evaluation outcome. Such consistent effects were never apparent. Repetition of a caveat is in place. Although the basis of the foregoing evidence is quantitative, we do not want to leave the impression that this was the sole or even most or very important argument. The practical value and meaningfulness of observed programme benefits were the focus. One reason for such a focus is that probability statistics and decisions may seem overly clear and strong, especially before having in detail studied and documented the effect of clustered sampling of learners into classrooms and schools, with its required adjustments to statistical significance outcomes. This task is seen as a future one, in academic forums and outputs, and not for the present report format.

*What appear to be the chances of and dynamics pertaining to scaling up so far?* The nature and design of the programme, the recruitment and training of its implementation staff, and indications of frequencies, volumes and materials relevant to the intervention, are all described briefly in the report. A single notional figure as to its cost per learner was also provided by sJsk. Alongside these, many indications of the dynamics underpinning the outcomes, and the greater effect of the programme on its recipients, compared to a control group, were provided. The foregoing elements document a strong likelihood that the observed impact can again be achieved by those following the same route. If one heeds the lessons learnt so far and is prepared to incorporate adjustments to current procedures as one goes, all bodes well. It also has to be said that school selection for Years 2 and 3 has been designed in such a way that this fourth question would be responded to best only once those stages have been embarked on and progressed for a while. That will bring to light how best a new layer of coordinators is brought on board, and how the impetus and quality of the interventions can be maintained. It would also bring to light how best to maintain the recruitment, training and allocation of a new group of facilitators. One specific challenge will be related to adding a second regional language, Tshivenda, to the intervention programme currently only having a Xitsonga footprint. This would double the implications related to recruiting local facilitators, doing the required training and producing the intervention materials. What is more, the progress of the first cohort to subsequent grades also requires the development of new intervention and assessment materials. In the process it is imperative that good communication with all stakeholders is maintained. Therefore, only with more information during and after 2014 will we be able to extrapolate from existing economies of scale relating to hierarchies of supervision, materials production, recruitment and training, etc. to the requirements and successes at new levels of functioning.

## 4.2 Implications and recommendations

Outcomes are very positive after one year over which sJsK's intervention implementation had run. It definitely appears possible to put language acquisition back on track and rebuild absent or damaged language foundations in a short time. This would be easier to accomplish and could happen quicker and more strongly at the earlier grade levels (Grade 1 to 4). It may require more than a full year at the level of Grade 7, though. The explanation for this would be the extent and depth of conceptual gaps that have already developed by then among learners.

It will only be possible to answer some remaining questions by continuing the evaluation for the remaining period. This means that the first cohorts have to be followed for two more years and the interventions have to be expanded as planned to the new cohorts and schools on an annual basis. *How much will learner achievement scores still improve in future? Will the tempo of improvement decelerate at some point? Can newly gained conceptual understandings and language content knowledge be sustained over time?* These are the kinds of questions that beg answers as the project unfolds further.

The intervention and evaluation models and sampling choices have so far combined well to provide meaningful answers to these and similar questions. What is required now is to continue learner achievement assessments into the second year, and to keep the intervention implementation records up-to-date. These should provide information on intervention volumes and any other dynamics related to intervention quality and systemic uptake. This would also ensure that complete second-year data are ready in a year so that a two-year impact analysis can be completed and reported on early in 2015 for the first cohort, and another one-year evaluation can be produced for the second cohort starting now.

On the topic of data, their availability and quality require ongoing diligence. A few comments are made here about this topic. One matter that was detected early on was early and strong coordination between the evaluation and intervention teams to ensure that learner background and achievement data can be linked more easily. However, attempts to get this right up-front are still being frustrated by the inability of schools to provide complete class lists in time. This is the result not only of poor infrastructure and capacity at schools for doing so, but also late movements of learners between schools into a new year. This is especially pertinent at the Grade 1 level, it seems.

It would also make much sense, perhaps only on a sample basis, that is, not necessarily by capturing all the records for every learner, to have item statistics available for each test. This would make two outcomes possible. First, it would enable calculations related to the psychometric properties of the tests. The project team would then know more about how consistently and reliably the items and tests measure the constructs of interest. At stake is the calculation of key internal consistency, re-test, reliability and correlation coefficients. Demonstrating instrument reliability, mainly through producing and using the appropriate item- to test-score matrices and correlations, serve as an important prerequisite for further claims about the quality of the instruments and their administration, and the validity of the assessments. It may even enable studying test or item bias. Second, it would enable deriving nuanced impressions of the nature and extent of existing conceptual gaps and shortcomings in learners' language proficiency and knowledge. This refers to knowing which learners struggle with which parts of the work. It would enable customising some of the thrusts of subsequent intervention steps, both in terms of the concepts and learners of focus, and in terms of monitoring adjusted intervention foci. This would be assessment for learning in action.

Pertaining to data still, a matter that was identified and hoped for already at the time of the provisional report produced after June 2013, is about smoothing data-collection operations at schools in relation to the contextual baseline data for new cohorts. Not only would that allow for more complete and efficient completion of questionnaires, but also easier linking of learner home and achievement data. This would rely strongly on generating better class lists earlier in the field. Such a resource would serve the purpose of

pre-numbering and auto-inserting unique learner record numbers to all instruments in advance. However, for the reasons stated in the previous few paragraphs, this goal seems to elude us all still.

Given the experience gained by all of those involved in this stimulating programme so far, it is anticipated that such foundations will strongly support honing and implementing the next phases from Year 2 onwards. It would in particular include all the processes, communications, records and data related to ending one year's evaluation cycle and starting up another. It would serve to keep all the beneficiaries, participants and other stakeholders on the same page, as it were. The latter is considered to include the relevant staff from the education department, especially at circuit level and in the schools, and also learners and their parents or caregivers.

High levels of excitement, gratitude and anticipation are very visible among virtually everyone involved in the programme thus far. This seems to indicate that the learner and language regeneration intervention has become institutionalised in the eastern Vhembe area to a real extent. It in turn predicts ongoing commitment to and involvement in the next phases of the work.

## Endnotes

---

<sup>i</sup> In relation to analysis of the **Grade 1** data:

- For the EGRA Letter Sound test, the value-added model explained 69,5% of the variance in the November scores, with the two standardised (beta) coefficients being 0,133 for the February score and 0,821 for intervention status (\*\*p<001). Adding monitoring status during February and November and their interaction terms with intervention status in a robustness check increased the variance explained by the model only marginally to 72,7%, and reduced the coefficient for intervention status to 0,528, remaining the largest contribution by far.
- For the EGRA Word Recognition test, the value-added model explained 53,2% of the variance in the November scores, with the two standardised (beta) coefficients being 0,046 for the February score and 0,726 for intervention status (\*\*p<001). As applied above, the robustness check increased the variance explained by the model only marginally to 56,3%, and reduced the coefficient for intervention status slightly to 0,682, remaining the largest contribution by far.
- For the EGRA Non-Word Decoding test, the value-added model explained 54,4% of the variance in the November scores, with the two standardised (beta) coefficients being -0,005 for the February score and 0,738 for intervention status (\*\*p<001). As applied above, the robustness check increased the variance explained by the model only marginally to 57,8%, and reduced the coefficient for intervention status slightly to 0,554, remaining the largest contribution by far.
- For the Oral Pictorial test, the value-added model explained 41,6% of the variance in the November scores, with the two standardised (beta) coefficients being 0,027 for the February score and 0,644 for intervention status (\*\*p<001). As applied above, the robustness check increased the variance explained by the model only to 49,3%, and actually increased the coefficient for intervention status to 0,768, remaining the largest contribution by far.

<sup>ii</sup> In relation to analysis of the **Grade 4** data:

- For the Schonell Reading Age test, the value-added model explained only 26,4% of the variance in the November scores, with the two standardised (beta) coefficients being 0,447 for the February score and 0,284 for intervention status (\*\*p<001). Adding monitoring status during February and November and their interaction terms with intervention status in a robustness check increased the variance explained by the model very marginally to 26,8%, and reduced the coefficient for intervention status to 0,208, remaining the second largest contribution behind the February coefficient now reduced a little to 0,436.
- For the EGRA Letter Sound test, the value-added model explained 83,5% of the variance in the November scores, with the two standardised (beta) coefficients being 0,028 for the February score and 0,906 for intervention status (\*\*p<001). As applied above, the robustness check increased the variance explained by the model only marginally to 84,6%, and reduced the coefficient for intervention status to 0,739, remaining the largest contribution by far.
- For the EGRA Word Recognition test, the value-added model explained 49,8% of the variance in the November scores, with the two standardised (beta) coefficients being 0,596 for the February score and 0,461

- 
- for intervention status (\*\*p<001). As applied above, the robustness check increased the variance explained by the model only marginally to 50,7%, and reduced the coefficient for intervention status slightly to 0,411, remaining the second largest contribution behind the February coefficient now increased a little to 0,610.
- For the EGRA Non-Word Decoding test, the value-added model explained 42,1% of the variance in the November scores, with the two standardised (beta) coefficients being 0,405 for the February score and 0,568 for intervention status (\*\*p<001). As applied above, the robustness check increased the variance explained by the model only marginally to 43,5%, and actually increased the coefficient for intervention status slightly to 0,573, remaining the largest contribution by a bit over the also increased February score coefficient of 0,422.
  - For the Literacy 1 (MCQ) sub-test, the value-added model explained only 4,9% of the variance in the November scores, with the two standardised (beta) coefficients being 0,136 for the February score and 0,152 for intervention status (\*\*p<001). As applied above, the robustness check increased the variance explained by the model to 12,5%, and decreased the coefficient for intervention status to 0,092, remaining one of the smaller contributions.
  - For the Literacy 2 (Sentence Ordering) sub-test, the value-added model explained only 3,0% of the variance in the November scores, with the two standardised (beta) coefficients being 0,026 for the February score and 0,172 for intervention status (\*\*p<001). As applied above, the robustness check increased the variance explained by the model to 12,3%, and also increased the coefficient for intervention status to 0,247, becoming one of three similar contributions.
  - For the Literacy 4 (Open-ended Items) sub-test, the value-added model explained only 12,2% of the variance in the November scores, with the two standardised (beta) coefficients being 0,293 for the February score and 0,202 for intervention status (\*\*p<001). As applied above, the robustness check increased the variance explained by the model to 18,3%, and decreased the coefficient for intervention status to 0,168, becoming the weakest contribution, also behind the February coefficient now reduced to 0,239.
  - For the Literacy Total overall test, the value-added model explained only 10,0% of the variance in the November scores, with the two standardised (beta) coefficients being 0,216 for the February score and 0,225 for intervention status (\*\*p<001). As applied above, the robustness check increased the variance explained by the model to 22,6%, and increased the coefficient for intervention status slightly to 0,234, becoming one of the middle-range contributions.

iii In relation to analysis of the Grade 7 data:

- For the Schonell Reading Age test, the value-added model explained 36,9% of the variance in the November scores, with the two standardised (beta) coefficients being 0,582 for the February score and 0,232 for intervention status (\*\*p<001). Adding monitoring status during February and November and their interaction terms with intervention status in a robustness check increased the variance explained by the model very marginally to 37,5%, and increased the coefficient for intervention status to 0,319, remaining the only other large contribution behind the February coefficient now reduced a little to 0,575.
- For the Literacy 1 (MCQ) sub-test, the value-added model explained only 7,8% of the variance in the November scores, with the two standardised (beta) coefficients being 0,164 for the February score and 0,237 for intervention status (\*\*p<001). As applied above, the robustness check increased the variance explained by the model to 13,1%, and decreased the coefficient for intervention status to 0,111, becoming the smallest contribution also behind the 0,138 coefficient for February's score.
- For the Literacy 3 (Paragraph Writing) sub-test, the value-added model explained only 12,3% of the variance in the November scores, with the two standardised (beta) coefficients being 0,116 for the February score and 0,342 for intervention status (\*\*p<001). As applied above, the robustness check increased the variance explained by the model slightly to 14,4%, and decreased the coefficient for intervention status to 0,254, becoming the strongest contributor by a small margin.
- For the Literacy 4 (Open-ended Items) sub-test, the value-added model explained 18,2% of the variance in the November scores, with the two standardised (beta) coefficients being 0,250 for the February score and 0,362 for intervention status (\*\*p<001). As applied above, the robustness check increased the variance explained by the model to 21,6%, and also the coefficient for intervention status much to 0,569, becoming by far the strongest contributor.
- For the Literacy Total overall test, the value-added model explained 26,0% of the variance in the November scores, with the two standardised (beta) coefficients being 0,349 for the February score and 0,411 for intervention status (\*\*p<001). As applied above, the robustness check increased the variance explained by the model to 29,4%, and the coefficient for intervention status slightly to 0,437, becoming the strongest contributor by a reasonable margin.

**Annexure 1: Comparative attrition (balance) statistics for Grade 1 learners: Paired and unpaired sub-groups**

Sub-test		Project group				Control group			
		February		November		February		November	
		Paired	Not paired	Paired	Not paired	Paired	Not paired	Paired	Not paired
Early Grade Reading Assessment (EGRA) Letter Sound	Mean	11,01 <sup>@</sup>	<sup>@</sup> 8,88	85,28 <sup>@</sup>	<sup>@</sup> 84,70	10,50 <sup>@</sup>	<sup>@</sup> 13,70	26,89 <sup>@</sup>	<sup>@</sup> 30,69
	(s.e.)	(0,87)	(0,68)	(0,85)	(0,86)	(1,00)	(1,39)	(1,36)	(1,84)
	SD	18,49	15,74	18,01	19,67	15,95	19,93	21,57	26,38
Early Grade Reading Assessment (EGRA) Word Recognition	Mean	0,03	0,04	61,92	59,98	0,01	0,07	11,79	13,08
	(s.e.)	(0,01)	(0,02)	(1,17)	(1,15)	(0,01)	(0,05)	(1,19)	(1,37)
	SD	0,23	0,49	24,60	26,52	0,13	0,70	18,97	19,55
Early Grade Reading Assessment (EGRA) Non-word Decoding	Mean	0,24	0,02	73,34	71,02	0,00	0,19	18,13	17,78
	(s.e.)	(0,22)	(0,01)	(1,10)	(1,14)	(0,00)	(0,16)	(1,65)	(1,73)
	SD	4,73	0,27	1,10	1,14	0,00	2,30	26,17	24,71
Oral Pictorial	Mean	13,59	14,20	90,09	88,92	12,66 <sup>@</sup>	<sup>@</sup> 15,12	55,32	56,50
	(s.e.)	(0,65)	(0,75)	(0,65)	(0,68)	(0,81)	(1,06)	(1,73)	(1,86)
	SD	13,85	17,27	13,81	15,59	12,86	15,13	27,45	26,68

\* p<0,05      \*\* p<0,01      \*\*\* p<0,001      (Correcting p-values after cluster sampling has not been done)

<sup>@</sup> Probability levels for these sets of comparison came within the 10% level (p<0,10)



Annexure 2: Comparative attrition (balance) statistics for Grade 4 learners: Paired and unpaired sub-groups

Sub-test		Project group				Control group			
		February		November		February		November	
		Paired	Not paired	Paired	Not paired	Paired	Not paired	Paired	Not paired
Schonell Reading Age	Mean	5,74*	*5,48	7,40*	*7,62	5,95	5,75	6,73	6,76
	(s.e.)	(0,07)	(0,10)	(0,06)	(0,09)	(0,07)	(0,16)	(0,06)	(0,10)
	SD	1,62	1,62	1,37	1,29	1,33	1,27	1,05	0,74
Early Grade Reading Assessment (EGRA) Letter Sound	Mean	25,05**	**19,22	92,26	93,32	11,87	10,40	17,66	18,81
	(s.e.)	(1,27)	(1,41)	(0,60)	(0,80)	(0,75)	(1,46)	(1,05)	(2,86)
	SD	28,38	22,57	13,55	11,93	14,16	11,42	19,72	21,22
Early Grade Reading Assessment (EGRA) Word Recognition	Mean	38,34	36,28	85,18	85,98	46,65*	*36,70	65,42	64,91
	(s.e.)	(1,44)	(1,92)	(0,87)	(1,30)	(1,65)	(4,13)	(1,50)	(3,69)
	SD	32,23	30,75	19,46	19,48	30,95	32,76	27,99	26,85
Early Grade Reading Assessment (EGRA) Non-word Decoding	Mean	39,44	40,32	88,87	89,42	48,48*	*38,79	63,52	65,50
	(s.e.)	(1,44)	(2,00)	(0,79)	(1,25)	(1,60)	(3,97)	(1,34)	(3,39)
	SD	32,15	32,09	17,66	18,68	30,05	31,53	25,11	24,43
Literacy 1: Multiple choice Items	Mean	21,48	20,55	39,29	38,65	12,49	10,90	28,08	26,00
	(s.e.)	(1,11)	(1,76)	(1,38)	(2,29)	(1,24)	(2,37)	(1,55)	(4,18)
	SD	25,39	26,11	31,71	32,99	23,45	17,11	29,28	29,58
Literacy 2: Sentence Ordering	Mean	20,21	21,10	38,26*	*32,54	23,38	18,63	27,68	28,50
	(s.e.)	(1,04)	(1,55)	(1,38)	(2,09)	(1,24)	(2,79)	(1,45)	(3,91)
	SD	23,97	22,94	31,76	29,89	23,38	19,90	27,23	27,67
Literacy 4: Open-ended Items	Mean	2,80	2,27	14,73	15,25	3,44	2,84	8,39	5,38
	(s.e.)	(0,32)	(0,38)	(0,68)	(1,02)	(0,45)	(0,81)	(0,89)	(2,04)
	SD	7,24	5,65	15,61	14,73	8,47	5,99	16,71	14,40
Literacy: Total Score	Mean	11,20	11,15	25,86	24,19	10,54	8,46	17,39	15,85
	(s.e.)	(0,42)	(0,68)	(0,78)	(1,08)	(0,55)	(1,17)	(0,89)	(2,20)
	SD	9,61	10,15	17,91	15,53	10,37	8,47	16,76	15,43

\* p<0,05

\*\* p<0,01

\*\*\* p<0,001

(Correcting p-values after cluster sampling has not been done)

Annexure 3: Comparative attrition (balance) statistics for Grade 7 learners: Paired and unpaired sub-groups

Sub-test		Project group				Control group			
		February		November		February		November	
		Paired	Not paired	Paired	Not paired	Paired	Not paired	Paired	Not paired
Schonell Reading Age	Mean	7,23**	**7,00	8,43	8,53	7,41	7,11	7,93***	***7,09
	(s.e.)	(0,04)	(0,09)	(0,06)	(0,14)	(0,06)	(0,28)	(0,06)	(0,23)
	SD	0,93	1,32	1,44	1,58	1,15	1,56	1,14	1,59
Literacy 1: Multiple choice Items	Mean	35,67*	*29,68	63,90	59,55	39,77	32,26	51,23	42,50
	(s.e.)	(1,19)	(2,25)	(1,00)	(2,19)	(1,61)	(5,24)	(1,59)	(4,46)
	SD	28,76	28,06	24,18	22,21	30,24	29,17	29,85	28,23
Literacy 3: Writing exercise	Mean	10,76	9,05	47,79	49,27	15,18	17,86	25,15	21,25
	(s.e.)	(0,87)	(1,39)	(1,42)	(3,35)	(1,17)	(3,79)	(1,37)	(4,25)
	SD	21,11	17,42	34,21	34,02	21,56	24,56	25,36	26,88
Literacy 4: Open-ended Items	Mean	4,89	3,39	27,50	27,43	6,63	7,06	11,40	11,00
	(s.e.)	(0,56)	(0,61)	(0,99)	(2,07)	(0,65)	(2,23)	(0,85)	(2,75)
	SD	13,50	7,62	23,79	20,99	12,16	12,98	15,89	17,36
Literacy: Total Score	Mean	15,30*	*12,39	42,23	40,80	18,28	17,02	26,66	22,50
	(s.e.)	(0,55)	(0,90)	(0,81)	(1,81)	(0,74)	(2,14)	(0,91)	(2,76)
	SD	13,24	11,21	19,69	18,36	13,93	11,91	16,99	17,43

\* p<0,05

\*\* p<0,01

\*\*\* p<0,001

(Correcting p-values after cluster sampling has not been done)

Annexure 4: Comparative monitoring statistics for Grade 1 learners: Monitored and not-monitored scores and gains for the project and control groups

Figures in each cell = Mean % (N) SD							
Monitored				Not monitored			
PROJECT SCHOOLS							
Sub-test	Monitor date	February	Feb → Nov	November	February	Feb → Nov	November
EGRA Letter Sound	February	15,7 (226) 22,0	70,7 (226) 26,2	86,3 (226) 17,4	***6,2 (221) 12,4	**77,9 (221) 20,8	84,2 (221) 18,6
	November	10,4 (364) 18,1	73,8 (364) 24,4	84,2 (364) 18,6	13,6 (83) 19,9	76,4 (83) 21,8	**90,0 (83) 14,1
	Either	9,2 (216) 15,6	75,1 (216) 21,3	84,2 (216) 18,0	**6,0 (44) 13,9	**83,6 (44) 19,4	89,6 (44) 16,6
	Both	14,3 (187) 21,7	71,2 (187) 27,1	85,5 (187) 18,3			
EGRA Word Recognition	February	0,1 (225) 0,3	64,8 (225) 25,6	64,8 (225) 25,6	*0,0 (220) 0,0	*58,9 (220) 23,2	*58,9 (220) 23,2
	November	0,0 (361) 0,3	59,2 (361) 24,9	59,3 (361) 24,9	0,0 (84) 0,0	***73,3 (84) 19,7	***73,3 (84) 19,7
	Either	0,0 (216) 0,0	60,6 (216) 22,7	60,6 (216) 22,7	*0,0 (44) 0,0	68,5 (44) 23,5	68,5 (44) 23,5
	Both	0,1 (185) 0,4	61,8 (185) 26,8	61,9 (185) 26,8			
EGRA Non-Word Decoding	February	0,5 (227) 6,6	76,3 (227) 24,7	76,8 (227) 23,6	0,0 (220) 0,0	**69,8 (220) 22,2	**69,8 (220) 22,2
	November	0,3 (363) 5,3	71,0 (363) 24,8	71,3 (363) 24,2	0,0 (84) 0,0	***82,1 (84) 15,6	***82,1 (84) 15,6
	Either	0,0 (216) 0,0	70,1 (216) 21,8	70,1 (216) 21,8	0,0 (44) 0,0	*80,8 (44) 18,1	**80,8 (44) 18,1
	Both	0,6 (187) 7,3	74,8 (187) 26,4	75,4 (187) 25,2			
Oral Pictorial	February	13,9 (227) 13,2	76,7 (227) 19,4	90,5 (227) 13,6	13,3 (222) 14,5	76,4 (222) 21,1	89,6 (222) 14,0
	November	14,3 (365) 14,2	74,8 (365) 21,0	89,2 (365) 14,5	*10,5 (84) 11,6	***83,7 (84) 14,5	**94,2 (84) 9,1
	Either	13,1 (218) 14,1	76,3 (218) 20,6	89,4 (218) 13,7	11,1 (44) 13,2	*83,4 (44) 16,7	94,5 (44) 10,2
	Both	14,8 (187) 13,7	75,1 (187) 20,4	89,8 (187) 14,5			
CONTROL SCHOOLS							
Sub-test	Monitor date	February	Feb → Nov	November	February	Feb → Nov	November
EGRA Letter Sound	February	6,8 (88) 10,6	18,0 (88) 19,8	24,8 (88) 19,3	**12,5 (165) 17,9	15,5 (165) 21,6	28,0 (165) 22,7
	November	9,7 (129) 14,8	11,0 (129) 18,4	20,8 (129) 14,5	11,3 (124) 17,0	***22,0 (124) 22,1	***33,3 (124) 25,6
	Either	8,5 (217) 13,3	13,9 (217) 19,2	22,4 (217) 16,7	***22,4 (36) 23,7	***31,6 (36) 24,6	***54,1 (36) 27,4
EGRA Word Recognition	February	0,0 (88) 0,2	10,6 (88) 19,2	10,6 (88) 19,3	0,0 (165) 0,0	12,4 (165) 18,8	12,4 (165) 18,8
	November	0,0 (129) 0,0	9,3 (129) 13,5	9,3 (129) 13,5	0,0 (124) 0,2	*14,4 (124) 23,1	*14,4 (124) 23,1
	Either	0,0 (217) 0,1	9,8 (217) 16,0	9,8 (217) 16,1	0,0 (36) 0,0	***23,6 (36) 28,9	***23,6 (36) 28,9
EGRA Non-Word Decoding	February	0,0 (88) 0,0	18,4 (88) 28,0	18,4 (88) 28,0	0,0 (164) 0,0	18,0 (164) 25,2	18,0 (164) 25,2
	November	0,0 (129) 0,0	12,8 (129) 17,9	12,8 (129) 17,9	0,0 (123) 0,0	**23,8 (123) 31,8	**23,8 (123) 31,8
	Either	0,0 (217) 0,0	15,1 (217) 22,7	15,1 (217) 22,7	0,0 (35) 0,0	***37,1 (35) 36,8	***37,1 (35) 36,8
Oral Pictorial	February	15,5 (88) 14,6	25,5 (88) 20,0	40,9 (88) 20,8	*11,2 (164) 11,6	***51,9 (164) 28,3	***63,0 (164) 27,5
	November	10,8 (128) 10,8	55,2 (128) 27,5	66,0 (128) 27,6	*14,6 (124) 14,4	***29,7 (124) 23,6	**44,3 (124) 22,6
	Either	12,7 (216) 12,7	43,1 (216) 28,7	55,8 (216) 27,9	12,5 (36) 14,0	40,0 (36) 28,4	52,5 (36) 25,0

\* p<0,05    \*\* p<0,01    \*\*\* p<0,001    \*\*\* = suggesting scoring bias; \*\*\* = counter-intuitive. (Correcting p-values after cluster sampling has not been done)

Annexure 5a: Comparative monitoring statistics for Grade 4 learners: Monitored and not-monitored scores and gains for the project and control groups

<u>Figures in each cell = Mean % (N) SD</u>		Monitored			Not monitored		
Sub-test	Monitor date	<b>(1<sup>st</sup> four subtests) PROJECT SCHOOLS</b>					
		February	Feb → Nov	November	February	Feb → Nov	November
Schonell Reading Age (age)	February	6,0 (275) 1,3	1,6 (275) 1,5	7,5 (275) 1,3	***5,5 (225) 1,9	1,8 (225) 1,7	*7,2 (225) 1,4
	November	5,8 (415) 1,5	1,6 (415) 1,6	7,4 (415) 1,3	**5,3 (85) 2,1	*2,0 (85) 1,6	7,3 (85) 1,8
	Either	5,5 (224) 1,8	1,8 (224) 1,8	7,3 (224) 1,1	***5,2 (43) 2,2	1,9 (43) 1,5	*7,1 (43) 2,3
	Both	6,1 (233) 1,1	1,5 (233) 1,5	7,6 (233) 1,4			
EGRA Letter Sound	February	30,3 (277) 31,6	61,6 (277) 33,5	91,9 (277) 14,9	***18,6 (225)22,3	***74,1 (225) 24,8	92,7 (225) 11,6
	November	26,3 (415) 29,5	66,9 (415) 31,0	93,2 (415) 11,6	*18,8 (87) 21,3	68,7 (87) 27,9	***87,6 (87) 19,9
	Either	19,4 (226) 22,6	71,0 (226) 26,7	90,4 (226) 14,3	***16,7 (43) 20,5	***76,9 (43) 22,8	*93,6 (43) 15,2
	Both	32,1 (233) 32,8	61,8 (233) 34,1	93,8 (233) 12,3			
EGRA Word Recognition	February	46,1 (276) 31,0	41,4 (276) 26,0	87,5 (276) 17,8	***28,8 (224)31,2	***53,5 (224) 27,5	**82,3 (224) 21,0
	November	37,6 (416) 32,3	48,4 (416) 27,7	86,0 (416) 18,0	42,1 (84) 31,8	**39,2 (84) 23,9	*81,3 (84) 25,3
	Either	28,6 (224) 30,9	54,5 (224) 27,0	83,1 (224) 18,2	***42,7 (42) 32,3	***36,6 (42) 23,7	**79,3 (42) 30,5
	Both	46,9 (234) 30,9	41,4 (234) 26,4	88,3 (234) 17,6			
EGRA Non-Word Decoding	February	48,2 (275) 29,8	42,0 (275) 29,0	90,2 (275) 16,1	***28,7 (224)31,7	***58,5 (224) 30,5	87,2 (224) 19,3
	November	38,0 (415) 32,3	51,5 (415) 31,2	89,6 (415) 16,1	*46,5 (84) 31,1	**39,0 (84) 26,1	85,5 (84) 23,6
	Either	28,5 (224) 30,4	59,7 (224) 28,9	88,1 (224) 15,7	***47,0 (42) 34,5	***36,0 (42) 28,5	*82,9 (42) 29,7
	Both	48,6 (233) 30,2	42,0 (233) 29,9	90,7 (233) 16,3			
		<b>(1<sup>st</sup> four subtests) CONTROL SCHOOLS</b>					
Sub-test	Monitor date	February	Feb → Nov	November	February	Feb → Nov	November
Schonell Reading Age (=age)	February	6,2 (113) 0,9	0,7 (113) 0,8	6,9 (113) 0,8	*5,8 (234) 1,5	0,8 (234) 1,3	6,7 (234) 1,2
	November	5,6 (149) 1,4	0,9 (149) 1,3	6,5 (149) 1,0	***6,2 (198) 1,2	0,7 (198) 1,0	***6,9 (198) 1,0
	Either	5,8 (262) 1,3	0,8 (262) 1,1	6,6 (262) 0,9	**6,3 (85) 1,5	0,7 (85) 1,3	**7,0 (85) 1,3
EGRA Letter Sound	February	9,5 (113) 11,8	7,5 (113) 21,2	17,0 (113) 17,2	*13,0 (239) 15,0	5,0 (239) 24,3	18,0 (239) 20,9
	November	11,8 (154) 12,4	0,8 (154) 18,3	12,6 (154) 14,8	11,9 (198) 15,4	***9,7 (198) 26,0	***21,6 (198) 22,1
	Either	10,8 (267) 12,2	3,6 (267) 19,9	14,4 (267) 15,9	*15,2 (85) 18,8	**12,6 (85) 31,2	**27,8 (85) 26,1
EGRA Word Recognition	February	54,6 (113) 29,6	15,5 (113) 15,9	70,1 (113) 26,1	**42,8 (237) 30,9	*20,3 (237) 21,4	*63,2 (237) 28,6
	November	36,9 (152) 28,5	23,9 (152) 21,7	60,8 (152) 26,6	***54,2 (198)30,7	***14,8 (198) 17,5	**68,9 (198) 28,6
	Either	44,4 (265) 30,2	20,3 (265) 19,8	64,8 (265) 26,7	*53,5 (85) 32,3	**13,9 (85) 19,4	67,4 (85) 31,7
EGRA Non-Word Decoding	February	56,0 (113) 27,1	13,0 (113) 23,6	68,9 (113) 23,0	**44,9 (238) 30,8	16,0 (238) 25,3	**60,9 (238) 25,7
	November	42,2 (153) 30,0	19,3 (153) 26,4	61,6 (153) 23,9	**53,3 (198) 29,2	**11,7 (198) 22,9	65,0 (198) 26,00
	Either	48,1 (266) 29,5	16,6 (266) 25,4	64,7 (266) 23,8	49,8 (85) 31,7	*10,1 (85) 21,9	59,9 (85) 28,8

\* p<0,05    \*\* p<0,01    \*\*\* p<0,001    \*\*\* = suggesting scoring bias; \*\*\* = counter-intuitive. (Correcting p-values after cluster sampling has not been done)

Annexure 5b: Comparative monitoring statistics for Grade 4 learners: Monitored and not-monitored scores and gains for the project and control groups

Figures in each cell = Mean % (N) SD							
Monitored				Not monitored			
(final four subtests) PROJECT SCHOOLS							
Sub-test	Monitor date	February	Feb → Nov	November	February	Feb → Nov	November
Literacy 1: MCQ	February	20,6 (286) 25,7	10,8 (286) 35,6	31,5 (286) 26,0	22,5 (240) 25,0	***26,1 (240) 39,6	***48,6 (240) 35,3
	November	22,5 (441) 25,9	16,5 (441) 38,9	39,0 (441) 31,3	*16,1 (85) 22,2	24,7 (85) 33,4	40,8 (85) 33,9
	Either	19,6 (243) 23,9	21,0 (243) 37,9	40,6 (243) 34,6	24,4 (41) 25,8	***39,8 (41) 35,1	***64,2 (41) 30,2
	Both	22,9 (242) 26,7	10,9 (242) 37,3	33,7 (242) 26,4			
Literacy 2: Sentence Ordering	February	20,9 (286) 24,7	8,6 (286) 32,8	29,5 (286) 25,2	19,4 (241) 23,2	***29,3 (241) 42,7	***48,7 (241) 35,4
	November	21,0 (442) 24,3	15,8 (442) 37,9	36,7 (442) 31,2	16,2 (85) 22,1	**30,0 (85) 42,7	*46,2 (85) 33,7
	Either	19,7 (244) 23,7	24,2 (244) 41,4	43,9 (244) 33,5	15,2 (41) 20,1	***45,1 (41) 43,0	***60,4 (41) 36,6
	Both	21,6 (242) 24,8	7,2 (242) 31,7	28,8 (242) 25,3			
Literacy 4: Open- ended Items	February	2,1 (286) 5,7	9,9 (286) 12,7	12,0 (286) 12,2	*3,7 (241) 8,6	**14,3 (241) 18,9	***18,0 (241) 18,4
	November	2,8 (442) 7,3	11,0 (442) 14,2	13,8 (442) 13,7	2,8 (85) 7,0	**16,9 (85) 22,6	**19,7 (85) 22,6
	Either	2,9 (244) 7,8	13,3 (244) 15,8	16,1 (244) 15,3	5,2 (41) 9,4	***20,1 (41) 28,7	***25,3 (41) 27,9
	Both	2,3 (242) 6,2	9,2 (242) 12,2	11,5 (242) 11,6			
Literacy: Total	February	10,8 (286) 9,2	9,7 (286) 14,1	20,5 (286) 12,1	11,7 (241) 10,0	***20,5 (241) 22,4	***32,2 (241) 21,3
	November	11,6 (442) 9,5	13,3 (442) 18,1	24,9 (442) 16,7	*9,0 (85) 9,8	***22,0 (85) 22,5	**31,00 (85) 22,5
	Either	10,7 (244) 9,8	17,6 (244) 20,3	28,3 (244) 19,0	11,7 (41) 10,6	***30,7 (41) 25,7	***42,4 (41) 25,5
	Both	11,6 (242) 9,2	9,0 (242) 13,8	20,5 (242) 12,2			
(final four subtests) CONTROL SCHOOLS							
Sub-test	Monitor date	February	Feb → Nov	November	February	Feb → Nov	November
Literacy 1: MCQ	February	12,3 (108) 22,6	12,0 (108) 32,0	24,4 (108) 25,6	12,6 (247) 23,9	17,1 (247) 35,3	29,7 (247) 30,7
	November	8,9 (157) 18,6	12,5 (157) 30,8	21,4 (157) 22,3	*15,3 (198) 26,4	18,0 (198) 36,9	***33,3 (198) 32,9
	Either	10,3 (265) 20,4	12,3 (265) 31,2	22,6 (265) 23,7	**18,9 (90) 30,0	**25,2 (90) 41,0	***44,1 (90) 37,3
Literacy 2: Sentence Ordering	February	23,4 (108) 22,0	-1,2 (108) 37,3	22,2 (108) 27,2	23,4 (247) 24,0	6,7 (247) 35,0	*30,1 (247) 27,0
	November	23,6 (157) 23,0	0,2 (157) 32,1	23,7 (157) 22,6	23,2 (198) 23,7	7,6 (198) 38,4	*30,8 (198) 30,1
	Either	23,5 (265) 22,6	-0,4 (265) 34,3	23,1 (265) 24,5	23,1 (90) 25,8	***18,1 (90) 37,1	**41,1 (90) 30,3
Literacy 4: Open- ended	February	3,3 (108) 6,9	2,0 (108) 14,3	5,3 (108) 12,4	3,5 (243) 9,1	*6,3 (243) 15,2	*9,8 (243) 18,2
	November	0,8 (154) 3,2	3,4 (154) 9,6	4,2 (154) 9,2	***5,5 (197) 10,5	6,2 (197) 18,1	***11,7 (197) 20,2
	Either	1,8 (262) 5,2	2,8 (262) 11,8	4,6 (262) 10,6	***8,1 (89) 13,2	***11,3 (89) 20,9	***19,5 (89) 24,7
Literacy: Total	February	10,5 (108) 9,0	3,1 (108) 14,8	13,6 (108) 13,1	10,6 (248) 10,9	**8,5 (248) 16,7	**19,0 (248) 17,9
	November	8,4 (158) 7,0	4,3 (158) 12,4	12,8 (158) 10,3	**12,2 (198) 12,2	**8,9 (198) 18,6	***21,1 (198) 19,8
	Either	9,3 (266) 7,9	3,8 (266) 13,4	13,1 (266) 11,5	***14,3 (90) 14,9	***15,7 (90) 20,5	***30,0 (90) 22,6

\* p<0,05    \*\* p<0,01    \*\*\* p<0,001    \*\*\* = suggesting scoring bias; \*\*\* = counter-intuitive. (Correcting p-values after cluster sampling has not been done)

**Annexure 6: Comparative monitoring statistics for Grade 7 learners: Monitored and not-monitored scores and gains for the project and control groups**

<b>Figures in each cell = Mean % (N) SD</b>		<b>Monitored</b>			<b>Not monitored</b>		
<b>PROJECT SCHOOLS</b>							
<b>Sub-test</b>	<b>Monitor date</b>	<b>February</b>	<b>Feb → Nov</b>	<b>November</b>	<b>February</b>	<b>Feb → Nov</b>	<b>November</b>
<b>Schonell Reading Age (age)</b>	<b>February</b>	7,3 (162) 0,8	1,2 (162) 1,4	8,5 (162) 1,6	7,2 (364) 1,0	1,2 (364) 1,1	8,4 (364) 1,4
	<b>November</b>	7,2 (453) 1,0	1,2 (453) 1,2	8,4 (453) 1,4	7,4 (73) 0,7	1,5 (73) 1,2	*8,8 (73) 1,4
	<b>Either</b>	7,2 (371) 1,0	1,2 (371) 1,2	8,4 (371) 1,4			
	<b>Both</b>	7,3 (122) 0,8	1,2 (122) 1,5	8,4 (122) 1,5	7,2 (33) 0,6	1,6 (33) 1,1	8,7 (33) 1,2
<b>Literacy 1: MCQ</b>	<b>February</b>	33,3 (161) 29,8	29,5 (161) 33,6	62,8 (161) 24,3	36,6 (423) 28,3	27,7 (423) 34,3	64,3 (423) 24,1
	<b>November</b>	33,9 (506) 27,9	28,6 (506) 34,3	62,5 (506) 24,0	***47,4 (78) 31,6	25,6 (78) 32,6	***73,1 (78) 23,5
	<b>Either</b>	36,5 (425) 28,5	28,1 (425) 34,4	64,5 (425) 24,8	**48,2 (38) 30,7	23,7 (38) 31,9	**71,9 (38) 19,8
	<b>Both</b>	28,9 (121) 27,5	30,2 (121) 33,6	59,1 (121) 22,4			
<b>Literacy 3: Writing exercise</b>	<b>February</b>	16,1 (161) 25,6	35,1 (161) 40,8	51,2 (161) 32,00	***8,7 (422) 18,7	37,8 (422) 38,2	46,5 (422) 35,0
	<b>November</b>	9,7 (505) 20,5	37,0 (505) 39,2	46,6 (505) 34,1	**17,9 (78) 23,5	37,5 (78) 37,3	*55,4 (78) 33,9
	<b>Either</b>	9,2 (424) 19,6	37,5 (424) 38,0	46,7 (424) 34,2			
	<b>Both</b>	14,8 (121) 25,5	34,8 (121) 42,3	49,6 (121) 32,8	*15,5 (38) 20,4	39,1 (38) 38,8	54,6 (38) 38,5
<b>Literacy 4: Open-ended</b>	<b>February</b>	6,1 (158) 9,8	21,2 (158) 20,8	27,3 (158) 21,4	4,4 (423) 14,6	23,1 (423) 25,3	27,6 (423) 24,7
	<b>November</b>	4,6 (503) 14,0	21,0 (503) 24,0	25,6 (503) 23,5	6,9 (78) 9,5	***33,1 (78) 22,9	***40,0 (78) 21,7
	<b>Either</b>	4,8 (425) 14,8	21,9 (425) 24,5	26,7 (425) 23,8			
	<b>Both</b>	5,1 (118) 9,5	19,5 (118) 20,9	24,6 (118) 21,5	4,6 (38) 8,3	***40,4 (38) 23,7	***45,0 (38) 23,5
<b>Literacy: Total</b>	<b>February</b>	16,3 (161) 11,9	26,2 (161) 16,9	42,5 (161) 17,8	14,9 (423) 13,7	27,2 (423) 20,1	42,1 (423) 20,4
	<b>November</b>	14,4 (506) 13,3	26,2 (506) 19,4	40,6 (506) 19,3	***21,3 (78) 11,6	*31,8 (78) 17,3	***53,0 (78) 18,8
	<b>Either</b>	15,2 (425) 13,9	26,6 (425) 19,6	41,8 (425) 19,9			
	<b>Both</b>	14,2 (121) 11,0	25,4 (121) 17,4	39,6 (121) 17,1	19,9 (38) 10,7	*35,1 (38) 18,8	***55,0 (38) 20,3
<b>CONTROL SCHOOLS</b>							
<b>Sub-test</b>	<b>Monitor date</b>	<b>February</b>	<b>Feb → Nov</b>	<b>November</b>	<b>February</b>	<b>Feb → Nov</b>	<b>November</b>
<b>Schonell Reading Age (age)</b>	<b>February</b>	7,5 (95) 0,9	0,6 (95) 0,8	8,1 (95) 1,1	7,4 (253) 1,2	0,5 (253) 0,9	7,9 (253) 1,1
	<b>November</b>	7,2 (163) 1,2	0,5 (163) 0,9	7,7 (163) 1,1	***7,6 (185) 1,0	0,5 (185) 0,9	***8,2 (185) 1,1
	<b>Either</b>	7,3 (258) 1,1	0,5 (258) 0,9	7,8 (258) 1,1	**7,8 (90) 1,1	0,5 (90) 0,9	**8,2 (90) 1,2
<b>Literacy 1: MCQ</b>	<b>February</b>	44,0 (103) 29,6	3,6 (103) 36,4	47,6 (103) 27,9	38,0 (249) 30,4	*14,7 (249) 39,8	52,7 (249) 30,6
	<b>November</b>	34,6 (159) 29,7	9,6 (159) 41,8	44,2 (159) 30,4	**44,0 (193) 30,1	13,0 (193) 36,8	***57,0 (193) 28,2
	<b>Either</b>	38,3 (262) 30,0	7,3 (262) 39,8	45,5 (262) 29,4	44,1 (90) 30,7	**23,7 (90) 34,3	***67,8 (90) 24,7
<b>Literacy 3: Writing exercise</b>	<b>February</b>	16,4 (103) 23,4	8,9 (103) 27,8	25,2 (103) 23,1	14,7 (238) 20,8	10,5 (238) 29,6	25,1 (238) 26,3
	<b>November</b>	12,8 (148) 20,4	5,2 (148) 30,0	18,1 (148) 25,1	17,0 (193) 22,3	**13,6 (193) 27,8	***30,6 (193) 24,3
	<b>Either</b>	14,3 (251) 21,7	6,7 (251) 29,1	21,0 (251) 24,5	17,6 (90) 21,2	**19,0 (90) 27,0	***36,7 (90) 24,3
<b>Literacy 4: Open-ended</b>	<b>February</b>	9,9 (101) 13,8	4,2 (101) 20,6	14,1 (101) 18,2	*5,3 (248) 11,2	5,0 (248) 14,9	*10,3 (248) 14,7
	<b>November</b>	3,8 (158) 9,3	3,4 (158) 11,2	7,3 (158) 11,0	***9,0 (191) 13,7	5,9 (191) 20,2	***14,8 (191) 18,3
	<b>Either</b>	6,2 (259) 11,6	3,7 (259) 15,5	9,9 (259) 14,6	7,9 (90) 13,5	7,7 (90) 19,7	**15,7 (90) 18,5
<b>Literacy: Total</b>	<b>February</b>	21,3 (103) 13,7	4,9 (103) 16,0	26,2 (103) 15,6	**17,0 (249) 13,8	*9,8 (249) 17,6	26,8 (249) 17,6
	<b>November</b>	14,9 (159) 12,9	7,0 (159) 17,0	21,9 (159) 16,2	***21,0 (193) 14,2	9,5 (193) 17,5	***30,5 (193) 16,7
	<b>Either</b>	17,4 (262) 13,5	6,2 (262) 16,6	23,6 (262) 16,1	20,7 (90) 14,8	***14,8 (90) 17,7	***35,5 (90) 16,5

\* p<0,05    \*\* p<0,01    \*\*\* p<0,001    \*\*\* = suggesting scoring bias; \*\*\* = counter-intuitive. (Correcting p-values after cluster sampling has not been done)