siyaJabula
siyaKhula
educates the nation

HSRC
Human Sciences
Research Council

kaMhinga Literacy Project

Hope for rebuilding language foundations: Part 2 - Replication in a second cohort ....Evidence after one year (2015)

Human Sciences Research Council Evaluation Report:

siyaJabula siyaKhula's Learner Regeneration Project
Vhumbedzi and Malamulele North-East, Vhembe, Limpopo
Evidence after one year (2014)

Harvey, J, Prinsloo, CH, Thaba, W & Moodley, M

September/October 2015

USAID
FROM THE AMERICAN PEOPLE

basic education
Department:
Basic Education
REPUBLIC OF SOUTH AFRICA

the ELMA foundation

J.P.Morgan

Table of Contents

1

List of Tables, Figures and Annexures

2

List of Tables, Figures and Annexures

3

List of Tables, Figures and Annexures

4

Foreword

Rather than a repetition of what was said in May last year in the first impact report (Prinsloo & Rogers, 2014) after completion of one year's learner regeneration interventions by siyaJabula siyaKhula (sJsK), we will reiterate the salient points and provide new points for consideration. Readers are also notified that the final impact report, on completion of the intervention period on 30 June 2015, is produced now in two parts. This outcome was determined by the intervention and evaluation design straddling two cohorts, with a second one introduced at the beginning of 2014 to run in parallel with the subsequent period following on the first year (2013) of the first cohort. The present report is confined solely to the second cohort. Essentially, this report is also a replication study to determine if the same results could be achieved a second time. Suffice to say at this stage that a deliberate attempt has been made to reduce the information that appears in this Part 2 replication report (Cohort 2, one-year) to an essential minimum. The companion report for the original group appears as the Part 1 report (Cohort 1, over two years) and will be the more comprehensive report.

The intervention was aimed at repairing critical gaps in learner's 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. The major change in this replication period was the scaling up of the number of schools involved and geographical area to straddle two circuits. The result of the scaling up means, amongst other implications, that an additional language community became involved. The study now had learners from Tshivenda-speaking homes and not, as before, only Xitsonga-speaking learners. The details of this are covered in the remaining pages.

The intended readership of this report and its companion version includes the funders (the United States Agency for International Development (USAID) and ELMA Foundation), its intervention agency (sJsK), the recipients or beneficiaries (officials and staff from the Limpopo and National Department of Basic Education, with its participating schools, circuits and district), the relevant local and parent communities, and academics. Beneficiaries are understood in particular to start with the learners, parents and teachers at the participating schools, and then also their school managers and management teams and governing bodies, not only for intervention schools, but also control schools. Everyone should have an opportunity to learn how the intervention achieved. We also now have the benefit of a second investigation, by means of a replication, essentially, of what seemed to be a very promising intervention after the first analysis last year.

We can emphasise the following few important points in relation to the contents and interpretation of the findings in this one-year Cohort 2 impact report:

- The promising intervention impact of the Cohort 1 first year was replicated.
- The impact was somewhat lower than anticipated.
- However, the above can be mitigated by understanding the influence of systemic events and factors outside the control of the intervention service provider.
- Using one sample of all Grade 7 learners from schools which had previously undergone the intervention and schools newly recruited for the second year did not introduce a new source of bias.
- Using only paired data, only using scores of those learners for whom baseline and end-of-year scores were available, did not introduce selection bias or unbalance the observed scores.

5

Foreword

- Analysing all data together, irrespective of whether a session was monitored or not, also did not introduce scoring or assessment bias as no systematic patterns were detected between monitored and unmonitored groups.

For more complete coverage of the evaluation methodology and instruments, the Year 1 report from 2013 and the Part 1 report should be consulted, as already indicated.

Last, it is confirmed that an external reader made extensive comments to a previous version of this manuscript. These have all been addressed. Another set of revisions then followed internal HSRC review and sign-off 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

Reference:
Prinsloo, C.H.& Rogers, S. (2014). Hope for rebuilding language foundations (HSRC Evaluation Report): siyaJabula siyaKhula's Learner Regeneration Project at Mhinga, Limpopo; Evidence after Year 1 (2013). Pretoria: Human Sciences Research Council.

6

Executive Summary

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.

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 and 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 Grade 1 and 4 in 16 project schools in 2014/2015 comprised the second cohort of the intervention population. All Grade 7 learners from the original 11 project schools and additional 16 project schools also comprised the second cohort[1]. Their anticipated achievement gains are compared with those of all learners from either four matched control-schools, for Grades 1 and 4, or from nine matched control-schools for Grade 7 learners. The evaluation data was 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 either Xitsonga- or Tshivenda-speaking. One would therefore with high confidence be able to attribute differing achievement

[1] In the second year, the intervention was enlarged to include additional schools. Grade 7 learners from the original schools from the first year of the intervention as well as from the additional schools were part of the second cohort.

7

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 final semester of 2014, 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 one of the more technical decoding proficiencies (Letter Sound) 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 273% and 312% (with relatively low baselines still explaining these high figures partly) on average for learners in project schools compared to improvements of 34% and 255% for learners in the control group on the Letter Sound and Oral Pictorial sub-tests respectively. Whereas beyond 70% 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 just over 80% of control-school learners achieved the 50% level in the case of the Oral Pictorial (Vocabulary) sub-test[2].

Grade 4 learners from project schools showed significantly higher gains compared to learners from control schools in all 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) and Comprehension (sentence ordering and open-ended responses). These findings represent improvements ranging from just over 20% to as high as 375% and once over 400% for learners from project schools compared to changes from a decrease of 42% to an increase of 95% for learners from the control group. Reading Age gains for learners in project schools were slightly less than double (1.3 years) that of learners in control schools (0.7 years). Only 5% of control-school learners were at their age-appropriate reading age compared to 32% of the project-school learners. For the Word Recognition and Non-Word Decoding sub-tests, the baseline percentage, between 64% and 78%, and the end-of-year percentage, over 90%, of learners from both project- and control-schools achieving at least 50% were not significantly different. However, 95% of project-school learners achieved that level in the Letter Sound sub-test whereas the percentage of control-school learners achieving 50% decreased to zero over the course of the intervention period. On the Letter Sound, proficiencies for both groups were below 10% at the baseline, indicating a large increase for the project-group learners. In relation to receptive comprehension (multiple-choice and sentence ordering items), almost 55 percentage-points more project-school learners compared to control-school learners achieved 50% or more on these two sub-tests (80% and 54% respectively compared to 25% and 29%). 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. These findings represent improvements ranging from just over 10% to 65% and over 500% for the EGRA Letter Sound sub-test for learners from project schools. This is in comparison to changes from a decrease of approximately 3% to an increase of 33% and over 150% for the EGRA Letter Sound sub-test for learners from the control group. Reading Age gains for learners in project schools (0.7) were several months greater than that for learners from control schools (0.2). With regards to age-appropriate reading age achievement, 3% of project-school learners achieved this

[2] 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.

8

level compared to 2% of control-school learners. It should be noted that these learners have to bridge a gap of almost four years of work. Similarly to the Grade 4 learners, for the Word Recognition and Non-Word Decoding sub-tests, the baseline percentage, between 88% and 90%, and the end-of-year percentage, over 90%, of learners from both project- and control-schools achieving at least 50% were not significantly different. However, 96% of project-school learners achieved that level in November 2014 in the Letter Sound sub-test compared to 20% of control-school learners. On the Letter Sound, proficiencies for both groups were below 10% at the baseline, indicating a large increase for the project-group learners. In relation to the aggregated literacy score, 22% of project-school learners achieved a score of at least 50% compared to 11% of the control-school learners. The corresponding end-of-year figures pertaining to the multiple-choice, paragraph writing and other written responses sub-tests were 100%, 40% and 7% respectively for project-school learners compared to the figures of 100%, 16%, and 3.5% for control-school learners.

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

This report includes results on the up-scaling of the intervention to include additional schools as a second cohort. The relevant information and findings for this second cohort replicate the positive impact on project-school learners noted in the initial report and further support the sJsK intervention and the consistency of the observed achievement gains. 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 8% to 20% of the assessment events at the baseline, 10% to 40% at the mid-year, and 30% during May 2015, were monitored depending on grade level and school sub-group, as well as practical and logistical factors related to setting up field-visit schedules. It should be noted that the Grade 7 learner assessments were not monitored as their assessments took place in November 2014 rather than May 2015 as the result of several factors explained in more detail in Section 3.2. Monitoring followed the same procedure as the initial report. 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. “Counterintuitive” refers to the fact that observed effects were actually contrary to what would have occurred had the service providers influenced assessment processes or marks in their favour 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 benefitted strongly in terms of the direct objectives of this 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 from this second cohort replicate and support those of the initial report and first cohort. Together, these results provide additional support to the positive impact of the sJsK intervention and the consistency of the observed achievement gains.

9

Introduction: The origins of the language intervention

On 17June 2011, the United States Agency for International Development (USAID) invited applications from local organisations involved in promising and innovative, as well as partly tested, approaches for building the capacity of school-based educators. The non-profit or community-based organisation, siyaJabula siyaKhula (sJsK), was already running a community-supported language learning and teaching intervention in Gauteng which 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. The HSRC was also involved in this area of inquiry and was performing a consortium study focusing on literacy teaching and learning in the Foundation Phase (Grades 1-3) classrooms. Due to the need for a formal impact assessment of the sJsK intervention, to evaluate its viability for local, provincial, and national implementation, sJsK and the HSRC pooled their efforts.

Following successful navigation through the two-step application process and the added support of another funding agency, the Elma Foundation, the intervention was formally implemented early in 2013. The proposed interventions, hereby evaluated, were put together and structured in response to the School Capacity & Innovation Programme (SCIP) call that USAID issued 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.Background

1.1 Literature

The initial literature as reflected in the Year 1 report highlighted the present crisis in literacy and language teaching and learning in schools and how sJsK's learner regeneration thrust promised to underpin a promising solution through learner, teacher and system capacitation.

1.2 The evaluation brief

Readers are referred to the Year 1 report for more details about: (a) the directions indicated by the call for intervention implementation and impact evaluation proposals; (b) the way in which key components from the joint sJsK-HSRC technical application proposed an evaluation rationale, plan and objectives in alignment with intervention contents and dynamics; and (c) the strategy accepted for both separating and aligning sufficiently independent evaluation (M&E) components with intervention implementation.

In summary, this two-part report reflects substantive refinement of both intervention and evaluation approaches and procedures, including data management and data analysis, 'tested', as it were, over two annual SCIP learning meetings, a first round of impact evaluation and report writing a year ago, and the experience of observing 24 to 30 months of intervention implementation. We have made a deliberate effort to give all evaluation contents and any resulting recommendations and policy indications a sense of finality in this finishing pair of reports too. We also have to reiterate that these two reports only cover the impact evaluation, making them henceforth focus on just the evaluation approach and methodology, and not the broader intervention programme.

10

2. Methodology and design

2.1 Evaluation purposes /objectives

The HSRC had to determine the impact of the intervention. It also had to study the influence of some circuit- and school-uptake factors and systemic issues on scaling-up an intervention, if indeed potentially successful.

As a result, the HSRC collected different types of baseline data and used certain procedures in adherence to requirements that would enable the evaluation to demonstrate impact. In this process many components were important, including: how indicators got identified, monitored and measured; certainty about the attribution of impact; best practises in handling data sources, collection frequencies and management; sensitivity to the concerns of programme beneficiaries and other stakeholders; keeping national and international strategic goals for improving reading skills in sight; maintaining data reliability, validity and effectiveness (in collection and analysis); and recording evaluation implications important for intervention replication at scale.

2.2 Approach to the study

A quasi-experimental experimental- and control-group design was employed and is described fully in the initial 2013 and Year 1 reports. However, it is necessary to reiterate that steps were taken to ensure that the control group of schools were similar in as many key respects as possible to the intervention schools. Therefore, a reasonably sound counterfactual situation resulted. This allowed for using statistical probability methods to determine what would have prevailed had interventions not taken place. Another artefact of the sampling strategy is the fact that the research participants (learners) belong to a relatively small number of intervention and control schools. Such clustering also influences the evaluation and analyses. In short, one should not rely too strongly on the quantitative information (i.e., statistically significant effects), but at least in interpreting them adjust / correct for the clustering, in that way balancing the seemingly very exact and certain nature of the statistical results with a more qualitative understanding of the value, meaning and usefulness of knowing which proficiency gains were possible among learners from intervention schools as opposed to learners from control schools. In so doing the focus is on the qualitative interpretation of meaningful effect sizes.

The statistical approach used to determine impact is known as double-difference or difference-in-difference analysis or comparison and is also described more fully in the first two reports.

2.3 Participants / sample

Readers are referred to the Year 1 report for more details about: (a) the choice of schools and area; and (b) the socioeconomic and contextual circumstances of the schools and villages chosen. Section 3.1 will again be devoted to summarise the so-called balance statistics in support (or refutation) of this claim for baseline similarity.

On selecting these village areas, a complete population of schools, teachers and learners, as it were, participated in the intervention implementation and evaluation. Readers are reminded that the original call envisaged a three-year study. It would involve a first cohort over three years, a second one over two years, and a third one over a one-year period, all fitted into 2012 to 2104. The delayed selection of service providers towards the end of the first semester of 2012 initially saw the start and finish points changed to mid-year (of 2012 and 2015 respectively). This soon got reduced to a 2½-year roll-out to allow service providers to use the second semester of 2012 as a ramp-up period for getting systems in place and recruit staff. However, after two

11

periods of school disruption because of the service-delivery protests early in 2014 and cash flow limitations at the end of 2014, the first semester of 2015 merely served to deliver the final parts of the normal Year-2 interventions from 2014, applicable to both of the only two cohorts remaining as a result. Some further explanations have been provided at the outset (see Section 1.1). The evaluation in the current part-report, therefore, is about the second cohort of learners in Grade 1, 4 and 7 in 2014 who reached the end of their (revised) one year of involvement. A single service provider, sJsK, was responsible for the complete intervention implementation. Other than the first cohort from 2013, who received a (revised) second year of interventions in 2014, this second cohort would not continue receiving further intervention implementation as Grade 2 and Grade 5 learners in 2015. The only other defining feature of this second cohort is the expansion to additional project and control schools from the existing Malamulele North-East Circuit and new Vhumbedzi Circuit. Interventions terminated during the beginning of May 2015.

The eventual sample realisation, or completion rates, is not reported here, but in the findings section (3.2) to avoid repetition. Suffice it to say that Cohort 2 interventions and evaluation since 2014 involved all Grade 1, 4 and 7 learners and their teachers from an additional 16 project and 4 control schools. Schools on average had about 50 learners per grade. Year groups of fewer than 30 to 40 learners were the exception, as were extremely large groups of over 120.

2.4 Instruments

Two main types of instruments were used in the evaluation. They are background or contextual questionnaires and a range of learner achievement assessment tools. These are fully described in the Year 1 report.

2.4.1 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.

From the perspective of the intervention service provider, regular assessment and immediate feedback had to enable targeting and shaping intervention contents and learning materials to learner needs. This process had to account for foundational gaps and subsequent growth trajectories of individual learners on a weekly to monthly basis within facilitation groups of 8 to 10 learners, and allow for differential pacing and customisation of learning materials and “lessons”. Semesterly or annual testing, with results delays, would be insufficient.

From the perspective of evaluation independence, because testing and scoring had to be done in the field regularly, 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 forms were completed on the basis of observations made by HSRC staff during all assessments (February 2014 baselines, June 2014 mid-years and May 2015 end-lines). Findings pertaining to these also appear towards the end of this report (Section 3.3).

2.5 Evaluation procedures

Evaluation procedures remain as described in the Year 1 report of 2014. In short, it entailed using the instruments and procedures briefly summarised next.

Establishing the contextual baseline at school, teacher/classroom and learner/home level was accomplished by full-time HSRC research staff and post-graduate interns at the end of 2013 (October /November) and the

12

beginning of 2014, before interventions started. This pertained to the new cohort of Grade 1, 4 and 7 learners in the study in 2014. Background information was collected by means of self-report questionnaires. However, an assisted process was followed at schools in the case of parents (including regular caregivers) for the provision of learner home context information.

Baseline, end-of-first-semester and end-of-year learner achievement assessments were administered by sJsK during February 2014, June 2014 and May 2015. The exception was the Grade 7 group which was assessed in October/November 2014. At the same time, the HSRC conducted a higher than usual portion of quality assurance visits (8% to 30%). This portion was much higher than the customary 10% sample to account for the joint obligations between sJsK and HSRC in this regard, as explained already. Coverage varied depending on time point of assessment, grade level and school status (project or control). 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). The standardised assessment administration manual that HSRC assisted sJsK to develop at the beginning remained in use throughout. Facilitators trained by sJsK continued to conduct all learner assessments accordingly. The HSRC's impact analyses are based on the data collected as described here.

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, three main sets of findings are covered. The discussion pertains to: contextual equivalence between project and control schools (3.1); the impact of the second year of intervention implementation (3.2); and comparing learner achievement marks on the basis of assessment sessions having been monitored or not. Uptake factors and systemic issues influencing roll-out and scaling up are only discussed in the Year 1 report.

3.1 Contextual background and its equivalence between project and control schools

Background information had been collected by February 2014 before interventions commenced. Similar analyses to those which were done for the first cohort to determine if the contexts of the project and control schools were performed for the second cohort. For this cohort, 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 was kept in mind during analysis.

The results were sufficiently equivalent to rule out the effect of bias when ascribing anticipated achievement gains among learners in project schools to the project interventions. In order to allow a wider audience access, an abbreviated version of the findings is included in this report. 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.

The purposes of collecting sufficient background information about schools, teachers (classrooms) and learner homes remain the same as for the first cohort, namely, structured information regarding the three-year programme; contextual equivalence; and fully understanding the context. The latter two components 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.

13

3. Evaluation findings

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 16 project and four control schools, with the exception of Grade 7 learners who cluster into 27 project and nine 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.

The contextual equivalence of the first cohort's 11 project and five control schools was reported on in the initial 2013 report. With regards to the second cohort, 16 project and four control schools were identified in advance and school equivalence for these schools is reported herewith.

Six questionnaires were completed and returned, five being from project schools. Voluntary participation is a cornerstone of respect to research participants. Due to the low completion rate, statistical analyses could not be performed and therefore what follows is a qualitative assessment of school-level contextual equivalence.

Staff capacity and responsibility

The majority of schools indicated that they had a full complement of teachers or had a single non-language teacher vacancy. All schools indicated a single Principal, the lack of a Deputy-Principal, and either the lack of or one HoD for language as well as for other subjects. Schools marginally differed (1-2) in the number of language teachers in FP, IP, and Grade 7. The control school which completed the questionnaire did not indicate the complement of other subject teachers or administration staff. Within the project schools there was much variation in the numbers of teachers for other subjects in the FP, IP, and Grade 7 but either a lack of or one administration staff member. The project schools indicated that the SMT members and language teachers attended training between two and four times, whereas the control school principal indicated 10 sessions of training were attended.

The control school indicated one class per grade with relatively small numbers of learners (21-36 learners per class). The project schools indicated one to two classes per grade with large learner numbers (40-67 learners per class).

14

In sum, the main areas of dissimilarity between the control school and project schools lies in the number of training sessions attended by the SMT members and language teachers and in the number of classes and learners per class. However, without additional completed questionnaires it is difficult to ascertain the significance of these differences.

Infrastructure, facilities and conditions

The control schools indicated average to good electricity, tapped water, permanent buildings, an office block for the Principal and administration staff, usable classrooms, and photocopier. None/very poor to poor was indicated for flushed toilets for staff, flushed toilets for learners, telephone/s, fence/security, hall, staff room, library, computer facilities for staff/office, computer centre/facilities for learners, language laboratory, storage/walk-in safe room, and sports grounds.

In project schools, the majority had no or a very poor office block for principal and administration staff, flushed toilets for staff, flushed toilets for learners, telephone/s, hall, staff room, library, computer centre/facilities for learners, language laboratory, and storage/walk-in safe room. Other facilities were generally rated as average or good.

A rather balanced distribution of conditions was evident. The two groups of schools can be considered 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. In addition, Cohort 2 schools largely resembled Cohort 1 schools.

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.

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 77 teacher records in the dataset, with 11 teachers from control schools and nine who did not indicate their project group status. This should be kept in mind during the discussion as it is a small sample size, particularly with regards to the control group teachers. The distribution in relation to the sex of teachers from project schools shows that 21 (37%) were male and 36 (63%) were female. The corresponding figures for teachers from control schools were two (18%) male and 9 (82%) female.

Teacher capacity and responsibility

Qualification levels and experience between project- and control-school teachers are summarised in Table 1. 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 a three-year degree. The distribution differed very little and not significantly between project- and control-school teachers.

Table 1: 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 (**20.1**>17.0)	Overall years teaching language at FP (9.1<**11.3**)
Overall years teaching language at IP (**8.4**>5.5)	Teaching FP language in present school (5.3<**10.4**)
Overall years teaching Gr 7 language (**8.5**>3.7)	Years teaching FP in present post (4.8<**10.4**)
Teaching IP language in present school (**7.3**>0.9)	
Teaching Gr 7 language in present school (**7.2**>0.0)	

15

Conditions rated more highly in project schools	Conditions rated more highly in control schools
Years teaching IP in present post (**5.9**>0.0) Years teaching Gr 7 in present post (**6.0**>0.0) FP = Foundation Phase IP = Intermediate Phase	
No substantive difference	
Highest qualification level[3] of teacher – (2.0<**2.5**)	

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 a third of the teachers completed at least a diploma; 34% of project school teachers and 18% of control school teachers. More teachers at project schools had three-year degrees (36% compared to 18% at control schools), while more control school teachers had four-year degrees (64% compared to 30% at project schools). None of the teachers had post-graduate qualifications. 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 Tshivenda. However, it should be noted that several of the control-group teachers did not list their major subject which made their sample very small. For the project group, the second most common major for their highest qualification was English. A majority of project-school teachers reported having studied English as a major in their next-highest qualification, compared to control-school teachers who again mostly reported Tshivenda.

Teaching loads and patterns were very similar between teachers from project and control schools. Most teachers taught three subjects.

Classroom infrastructure, facilities and materials

The existence and quality of classroom infrastructure and facilities, rated on a five-point Likert scale[4] , are summarised in Table 2. Teachers from project and control schools reported working under largely similar conditions in their classrooms. Electricity, lighting, air quality, and comfortable temperatures were generally rated as very good. Chairs and desks for learners, overall space, permanent building materials, and chairs and desks for teachers were rated as around average to good. Lockable and other storage was rated as average whilst shelving was considered to be in shortage and rated as poor. There were no statistically significant differences across a range of probabilities in relation to the quality ratings between the two teacher groups.

Table 2: 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
Learner desks (**3.8**>3.6) Shelves (**2.2**>1.7) Comfortable temperature (**4.5**>3.8) Permanent building materials (**3.8**>3.7) Teacher chair (**3.7**>3.6) Lockable and other storage (**3.1**>3.0)	Electricity (4.0<**4.3**) Enough fresh air (4.0<**4.2**) Learner chairs (3.4<**3.9**) Overall space (3.6<**4.0**) Teacher desk (3.5<**3.8**)
No difference	
Appropriate light / enough windows (4.2~4.2)	

[3] 0=Gr 12; 1=Diploma; 2=Degree; 3=Honours / four-year degree; 4=Master's or doctoral degree
[4] 1=None/very poor; 2=Poor; 3=Average; 4=Good; 5=Very good

16

The availability of specific materials in classrooms, and their condition, rated on a five-point Likert scale[5], are summarised in Table 3. There were no differences with regard to provision of materials. Although quality was always rated higher at control schools, many such differences were rather small and not significant. Even significant differences are of little practical meaning, and would fall away on correction for clustering in the sample design.

Table 3: Availability and quality of materials used in classrooms for project and control schools

Conditions rated more highly in project schools	Conditions rated more highly in control schools
	Chalkboards (3.6<**4.3**)
	Chalk (3.7<**4.3**)
	Prestik/adhesives (3.0<**4.0**)
	Wall charts (2.8<**3.0**)
	Board erasers (3.4<**4.3**)
	Drawing pins (2.5<**4.0**)*
	Rulers or T-pieces (3.1<**4.3**)
No difference	
Availability of chalkboards, chalk, board erasers, drawing pins rulers or T-pieces (All schools ~ All schools)	

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

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 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. In addition, Cohort 2 teacher profiles largely resembled those from Cohort 1. Qualification levels for Cohort 2 teachers may be slightly higher than in 2013, as would be the availability and condition of some classroom materials.

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 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 16 project and four 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 1 829 records. A few parents (n=226) did not complete the field for the sex of learners. Of the remaining 1 603 learners, 50% and 50.1% respectively for project and control schools were boys. Thus, the distribution was very similar across type of school.

[5] 1=Not applicable/none; 2=Poor; 3=Average; 4=Good; 5=Excellent

17

3. Evaluation findings

Parental characteristics

- Duration of family living in their village: 33.0 years for the project group; 25.1 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): This could not be calculated due to the answer variation, at times simply incorrect, captured on the questionnaire. This should, however, be inconsequential.

- Age of parent/s or caregiver/s: Almost two-thirds were 40 years of age or below (including almost 5% below 20 years in age), around 20% up to 50 years of age, almost 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 for below 20 years of age, 31-40 years of age, and older than 51 years (1-3 %-points). However, in the age groups of between 20 and 30 years of age as well as between 41 and 50 years of age, the project school learners' parents (caregivers) had higher percentages (6-8 %-points).

- Highest qualification of parents or caregivers: No schooling for approximately 7%, primary school for around 19%, Grade 8 to 9 16%, Grade 10 to 11 for approximately 27%, matric for around 21%, and any post-school qualification for almost 9%. Slightly greater proportions of parents / caregivers of children from control schools (1-3 %-points) were better qualified from primary school until Grade 9 whereas the opposite was true (1-4 %-points) for either no formal schooling or from Grade 10 onwards.

- Earning capacity of parents / caregivers: No income for 15%; social grants for 71%; salaries for 10%; an income from farming and from business or trade for 1% and 3% respectively. Slightly more of those caring for children from control schools had no income (3 %-points) or received social grants (5 %-points), while more of those caring for children from project schools earned their own salaries (7 %-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 38.6%; Tshivenda for 60.3%. Comment: A higher percentage of families of project-school learners spoke Tshivenda (69.7%) than families of control-school learners (30.1%). Correspondingly, families of control-school learners more often spoke Xitsonga (68.4% as opposed to 29.4% of project-school families). This should be inconsequential.
- Language spoken by learner at home: Xitsonga for 39.1%; Tshivenda for 60%. Comment: As above.
- Learner proficiency[6] in the home language:
    - Speaking: 3.5 in project group; 3.7 in control group. Comment: Xitsonga/Tshivenda is spoken very well.
    - Reading: 3.4 in project group; 3.4 in control group. Comment: Xitsonga/Tshivenda is read well.
    - Writing: 3.4 in project group; 3.4 in control group. Comment: Xitsonga/Tshivenda is written well.
- Parents'/caregivers' proficiency in the learners' home language:
    - Speaking: 3.5 in project group; 3.6 in control group. Comment: Xitsonga/Tshivenda is spoken very well.
    - Reading: 3.3 in project group; 3.1 in control group. Comment: Xitsonga/Tshivenda is read well.
    - Writing: 3.3 in project group; 3.1 in control group. Comment: Xitsonga/Tshivenda is written well.
- How often[7] learners speak English at home: 2.5 in project group; 2.3 in control group (*p<05).
- Learner proficiency in English (Likert rating scales are as above from here onwards):
    - Speaking: 2.4 in project group; 2.4 in control group.
    - Reading: 2.6 in project group; 2.6 in control group.

[6] 1=Not at all well; 2=Only a little; 3=Well enough; 4=Very well
[7] 1=Never; 2=Rarely; 3=Sometimes; 4=A lot / often

18

  - ° Writing: 2.6 in project group; 2.6 in control group.
- How often parents / caregivers speak English at home: 2.4 in project group; 2.3 in control group (*p<05).
- Parent / caregiver proficiency in English:
  - ° Speaking: 2.5 in project group; 2.7 in control group.
  - ° Reading: 2.7 in project group; 2.8 in control group.
  - ° Writing: 2.7 in project group; 2.8 in control group.

Overall comment: The differences above are mostly marginal and assumed to be of little practical consequence. English, Sepedi, Setswana, Xhosa, and Zulu 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 in some areas and for project-school learners in others. Similarly, for the ratings of carers' reading and writing proficiencies compared to those of the learners. Learners and parents / caregivers were reported to rarely or sometimes 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[8] learners have to travel to school: 1.6 (scale points) in project group; 1.8 in control group (***p<001). Over 80% of learners stay within three kilometres from school. The average would be between two and three kilometres. Fewer than 3% have to travel more than six kilometres. Comment: More project-school learners stay within 1km of school (14 %-points) whereas more control-school learners stay between 1 and 3km (7 %-points) and between 3 and 6km (7 %-points).

- How learners get to school: By far most learners walk to school (95%). Those not walking get taken to school by car (3%), or go by bus or taxi (both approximately 1%). Very few go by bicycle (< 0.5%). Comment: There were minimal differences in %-points between project- and control-school learners for each mode of transport (1-2 %-points).

- How long it takes learners to get to school: Almost 38% of learners get to school within 10 minutes, while an additional 53% get there within half an hour. Less than 1% 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 4 of how often learner homes were reported to have the listed items.

Table 4: Percentage of parents / caregivers indicating that leamer homes have the listed items

Item	Group		Item	Group		Item	Group	
	Project	Control		Project	Control		Project	Control
***Tap water	26.2	45.2	*Radio	83.1	77.3	***Satellite dish	40.6	29.4
***Electricity	95.5	81.7	Computer	15.4	13.8	***Television	88.2	80.4
Flush toilet	7.0	7.5	Laptop	11.8	15.0	*CD/DVD player	77.8	72.3
***Fridge	86.5	72.9	*Internet	8.2	12.4	Car	25.9	23.4
***Telephone	12.7	25.2	Cellphone	91.2	90.4	Bicycle	23.3	20.4

* 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 control-school homes had tapped water; less

[8] 1=Less than 1 km; 2=1-3 km; 3=3-6 km; 4=More than 6 km

19

than a third in the case of project-school learners' homes. The remaining commodities were all in low supply, with favourability either with the project- or control-school. The inconsistent distribution may rule out any advantage for either sub-group.

Exposure to language-related materials and behaviours

- Learner has access to a public library: 16% in project group; 4.2% in control group (***p<001).
- Learner has access to a dictionary at home: 43.7% in project group; 32.9% in control group (***p<001).
- How often[9] learner reads the following materials at home:
    - Books: 2.6 in project group; 2.6 in control group. Comment: Half of learners in each group never or almost never, and once or twice a week read books. Daily reading was reported in 29.5% of learners.
    - Magazines: 1.5 in project group; 1.6 in control group (*p<05). Comment: 65% of learners in both groups were said never or almost never to read magazines. Only about 4% read daily in the project-school group compared to nearly double the percentage points (7.6%) in the control-group.
    - Newspapers: 1.5 in project group; 1.5 in control group. Comment: 67.5% of project-school learners and 65% of control-school learners were reported never or almost never to read newspapers. Only about 4.5% of project-school learners and 7.6% of control-group learners read daily.

- How often[10] learner is given help by at least one parent / caregiver in relation to the following:
    - Homework: 3.3 in project group; 3.0 in control group (***p<001). Comment: Just over half the control-school learners and 63% of project-school learners reportedly were helped on a weekly basis or more often. Thirty percent received only incidental help once a month or less.
    - Information collection for home- or schoolwork: 3.0 in project group; 2.8 in control group (*p<05). Comment: 41.4% of control-school learners and 47.1% of project-school learners were helped on a weekly basis or more often. 36.2% obtained information once a month or less.
    - Homework completion checked: 3.3 in project group; 3.1 in control group (**p<01). Comment: 59.1% for control-school learners and 66.2% for project-school learners on a weekly basis or more often. Just over a quarter of learners' homework was checked once a month or less.
    - Preparation for tests / exams: 3.1 in project group; 3.0 in control group (*p<05). Comment: Almost 50% for control-school learners and 55.5% for project-school learners on a weekly basis or more often. Help was rendered once a month or less with test / exam preparation 30% of cases.
    - By paying someone else to give such help: 1.9 in project group; 1.6 in control group (***p<001). Comment: 14.6% of control-school learners and 21.2% of project-school learners were afforded this help on a weekly basis or more often. Such help was never or almost never secured in almost 60% of cases.
    - Ensuring learner gets to and from school: 3.4 in project group; 2.8 in control group (***p<001). Comment: 48.3% for control-school learners and 70.8% for project-school learners on a weekly basis or more often. This was ensured in almost 30% of cases once a month or less.
    - Looking at learner results to improve them: 3.3 in project group; 2.9 in control group (***p<001). Comment: 47% for control-school learners and 62.3% for project-school learners on a weekly basis or more often. Results were looked at once a month or less in almost 30% of cases.
    - Encouragement to read and write: 3.4 in project group; 3.1 in control group (***p<001). Comment: 58% for control-school learners and 69% for project-school learners on a weekly basis or more often. Encouragement was given once a month or less in 7%-16% of cases.
    - Listening to learner reading: 3.3 in project group; 3.0 in control group (***p<001). Comment: 49% for control-school learners and 59.2% for project-school learners on a weekly basis or more often. Learner reading was listened to once a month or less in 10%-26% of cases.
    - Asking about what learners had read: 3.2 in project group; 2.8 in control group (***p<001). Comment: 44.6% for control-school learners and 57,8% for project-school learners on a weekly basis or more often. Learner reading was asked about once a month or less in 14%-29% of cases.

[9] 1=Never / almost never; 2=1-2 times a week; 3=3-5 times a week; 4=Almost daily / daily
[10] 1=Never / almost never; 2=Once every 3-6 weeks; 3=Once every1-3 weeks; 4=Every 1-4 days

20

- ° Asking about learner schoolwork: 3.3 in project group; 2.9 in control group (***p<001). Comment: 52% for control-school learners and 65% for project-school learners on a weekly basis or more often. Schoolwork was asked about once a month or less in 11%-26% of cases.

- For how long[11] learner is assisted by parent giving daily homework assistance: 3.0 in project group; 3.1 in control group. Comment: 26% of learners received as much help as needed; 9% 1-2 hours; 18% half an hour to an hour; almost 47% less than an hour, including 10% receiving less than 10 minutes.

- How many[12] books of his/her own a learner has at home: 2.0 in both the project and control groups. Comment: Almost half were reported to have no such books; another almost 30% had fewer than five; about 17% had five to ten; fewer than 13% had more than 10 books, including 3% having more than 25.

- How many[13] other books there are at home: 1.9 in project group; 1.7 in control group (*p<05). Comment: Almost half were reported to have no such books; another almost 40% had fewer than ten; about 14% had 10-25; fewer than 6% had more than 26 books, including 2% having more than 100.

- How much time[14] a learner spends or is required to spend at home on a typical weekday on:
  - ° Work in the home language subject: 2.8 in project group; 2.6 in control group (*p<05). Comment: 15%-45% spent 20 minutes or less; 31% spent more than 40 minutes.
  - ° English work: 2.6 in project group; 2.6 in control group. Comment: 58%-65% spent 20 minutes or less; about 25% spent more than 40 minutes.
  - ° Work on other school subjects: 2.7 in project group; 2.7 in control group. Comment: 14%-42% spent 20 minutes or less; 25%-27% spent more than 40 minutes.
  - ° Watching television: 2.6 in project group; 2.4 in control group (*p<05). Comment: 22%-48% spent 20 minutes or less; 24%-28% spent more than 40 minutes.
  - ° Reading for pleasure: 2.3 in project group; 2.5 in control group (*p<05). Comment: two-thirds spent 20 minutes or less; 16% spent more than 40 minutes.
  - ° Playing with / visiting friends: 2.9 in project group; 2.8 in control group. Comment: 25%-40% spent 20 minutes or less; 40%-50% spent more than 40 minutes.
  - ° Buying food/groceries for the family: 2.1 in project group; 2.5 in control group (***p<001). Comment: 61%-78% spent 20 minutes or less; 12%-18% spent more than 40 minutes.
  - ° Doing home tasks (e.g., washing, gardening): 2.3 in project group; 2.5 in control group (**p<01). Comment: 62%-73% spent 20 minutes or less; 12%-15% spent more than 40 minutes.
  - ° Doing own paid part-time work: 1.9 in project group; 2.2 in control group (***p<001). Comment: 72%-75% spent 20 minutes or less; 15%-17% spent more than 40 minutes.

General comment: Project-group learners were consistently reported to receive slightly higher levels of exposure or support than control-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. Again, this situation in 2014/2015 is largely comparable to that of 2013. If anything, the exposure of both learners and parents to and practicing in English were rated a little higher now. So was access to digital technologies.

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

[11] 1=Very little (<10 minutes); 2=10-30 minutes; 3=31-60 minutes; 4=1-2 hours; 5=As much s/he needs
[12] 1=None; 2=Fewer than 5; 3=5-10; 4=11-25; 5=More than 25
[13] 1=None; 2=Fewer than 10; 3=10-25; 4=26-100; 5=More than 100
[14] 1=0-10 minutes; 2=11-20 minutes; 3=21-40 minutes; 4=More

21

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 took place. As a result, the different learner achievement gain trajectories can ultimately be ascribed to the sJsK intervention. Two main comparisons are made in relation to learner's assessment score changes from the baseline scores of February 2014 to the end-of-year scores of either November 2014, for the Grade 7 learners, or of May 2015 in the various tests at the other grade levels. Indirect comparison investigates the growing proportion of learners who achieved at least 50% in each test across learners from project and control schools.

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 further down 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.

Instructor selection and training

This remained the same as in the first year report.

Teacher training

All teachers at the programme schools received training. Teachers from the first cohort received 18 hours of training in literacy teaching, how learners learn, and how to modify the curriculum to reach the level of the learners. In comparison, teachers from the second cohort only received 8 hours of training. This was due to various interplaying factors. Initially, late payment of funds by the ELMA Foundation prevented the programme from operating during the final months of 2014. In addition, the situation was worsened by the civil unrest in the Malamulele area, which resulted in the continued suspension of operations until March 2015. It was at this point that sJsK chose to forego the remainder of teacher training as a result of limited time and funding remaining and the general state of disruption within the schools.

The intervention content

Readers are referred to the Year 1 report for more details about: (a) the level and amount of interaction with each learner according to their grade; (b) how the community member instructors delivered the intervention programme; (c) the materials supplied to each learner according to their grade; (d) the breakdown of each one-hour intervention for each learner according to their grade; and (e) the average cost per learner of the intervention.

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 comprised that information.

22

3. Evaluation findings

When reviewing the results, there are contextual factors which need to be taken into account. Towards the end of 2014, and well into 2015, the educational system in South Africa became turbulent and learners were unable to attend school regularly. The first six weeks of school were lost in 2015 as a result. Therefore, Grade 1 and Grade 4 learners completed their third assessment in May 2015, rather than November 2014 as was scheduled. The lost intervention delivery during this period was caught up in the period February 2015 to early May 2015. Grade 7 learners were assessed in November 2014 as they change schools following completion of their school-year and would therefore be inaccessible in May 2015. Due to the alternate assessment date two additional factors need to be taken into account when reviewing the results from these assessments and the general trends for Grades 1 and 4 over the course of the intervention.

Previous research has indicated the negative effect of a long school holiday period on learner achievement where learners achieve lower results on standardised tests following a three month holiday period (Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996; Rambo-Hernandez & McCoach, 2015). The “summer learning loss” effect can have an even stronger impact for children from families with a different home language from that of instruction (Cooper et al., 1996) , or from a low socioeconomic status (Alexander, Entwisle, & Olson, 2001; Cooper et al., 1996; Entwisle & Alexander, 1992).

In addition, maintenance at home of the achievement growth students acquired during the school year is difficult. Although reading and verbal skills can be acquired and improved on during both the school year and at home, as opposed to other skills such as mathematics which are mostly developed solely at school (Rambo-Hernandez & Mc Coach, 2015), learners' parents may be illiterate themselves and are therefore unable to effectively assist their children. Furthermore, the availability of books within homes is limited which prevents children from practicing their skills during the school holiday period.

Due to these factors, the May 2015 results for Grade 1 and 4 learners may be lower than the general trend calculated from baseline and mid-year scores would suggest, and in comparison to the February 2013 to November 2013 results of the first cohort whose intervention year was not disrupted. However, it is proposed that regular schooling would counter the summer learning loss and students would resume their general trend sometime into a new academic year.

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:

- 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 684 to 689 learners depending on the sub-test. Of these learners, 148 or 149 were from control schools and between 535 and 540 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

23

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. 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 5. 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 significant differences between learners at the outset, with control-school learners outperforming project-school learners in the EGRA Letter Sound sub-test, 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 2014 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 1 to 4. Where zero baselines do not render calculating improvement percentages meaningless, these findings represent improvements of 273% and 312% on average for learners in project schools compared to improvements of 34% and 255% 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.

Table 5: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 2 % (& SD)	Change % pts (& SD)	Diff-in-diff % pts (& SD)	Change % (& SD)
EGRA Letter Sound	Project	17.3 (17.1)	***64.5 (24.4)	***47.2 (26.2)	39.4 (2.3)	272.8(2.8)
	Control	***22.8 (16.8)	30.6 (17.4)	7.8 (19.4)		34.2 (0.5)
EGRA Word Recognition	Project	0.2 (1.8)	**39.0 (31.0)	**38.8 (30.8)	9.2 (@)	@(@)
	Control	0.17 (2.0)	29.7 (26.0)	29.5 (25.6)		@ (14.8)
EGRA Non-Word Decoding	Project	0.5 (3.3)	*50.1 (33.8)	*49.6 (33.5)	7.3 (8.7)	@(15.0)
	Control	0.8 (6.7)	43.0 (28.9)	42.3 (28.5)		@(6.3)
Oral Pictorial	Project	18.4 (17.5)	***75.8 (20.5)	***57.4 (23.6)	11.4 (0.2)	312.0 (3.3)
	Control	18.0 (14.9)	64.1 (21.2)	46.0 (23.3)		256.0(3.1)

* 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

24

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.

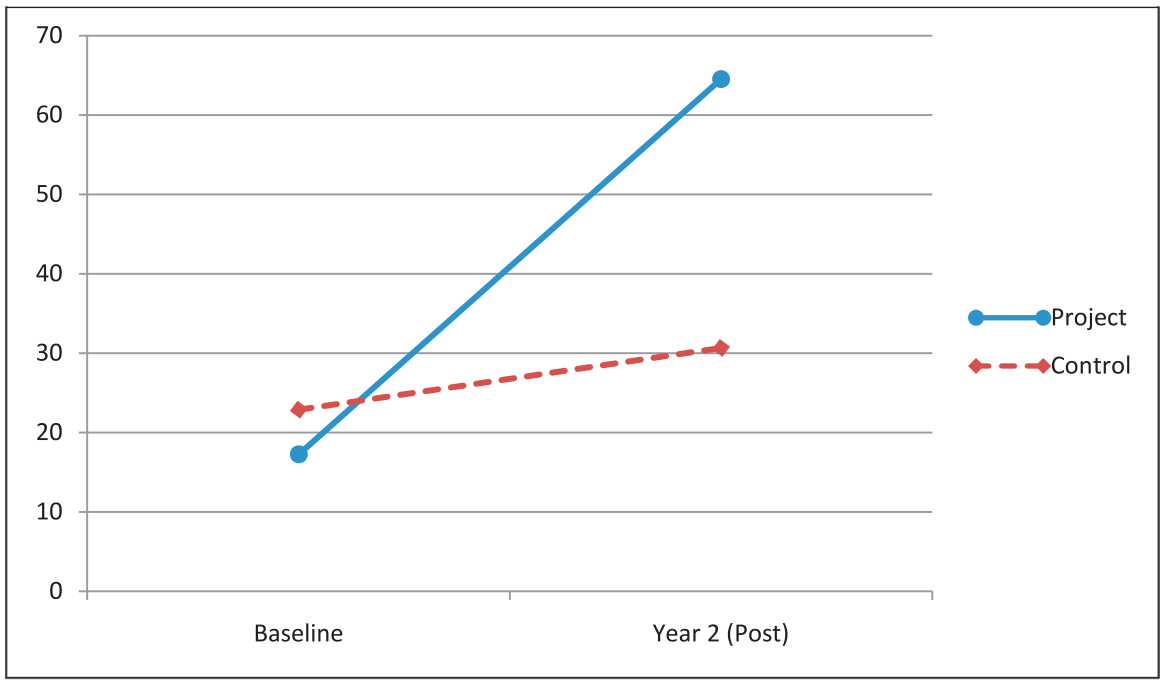

Figure 1: Grade 1 mean learner scores (%s) on the EGRA Letter Sound sub-test over time by school type (for Cohort 2)

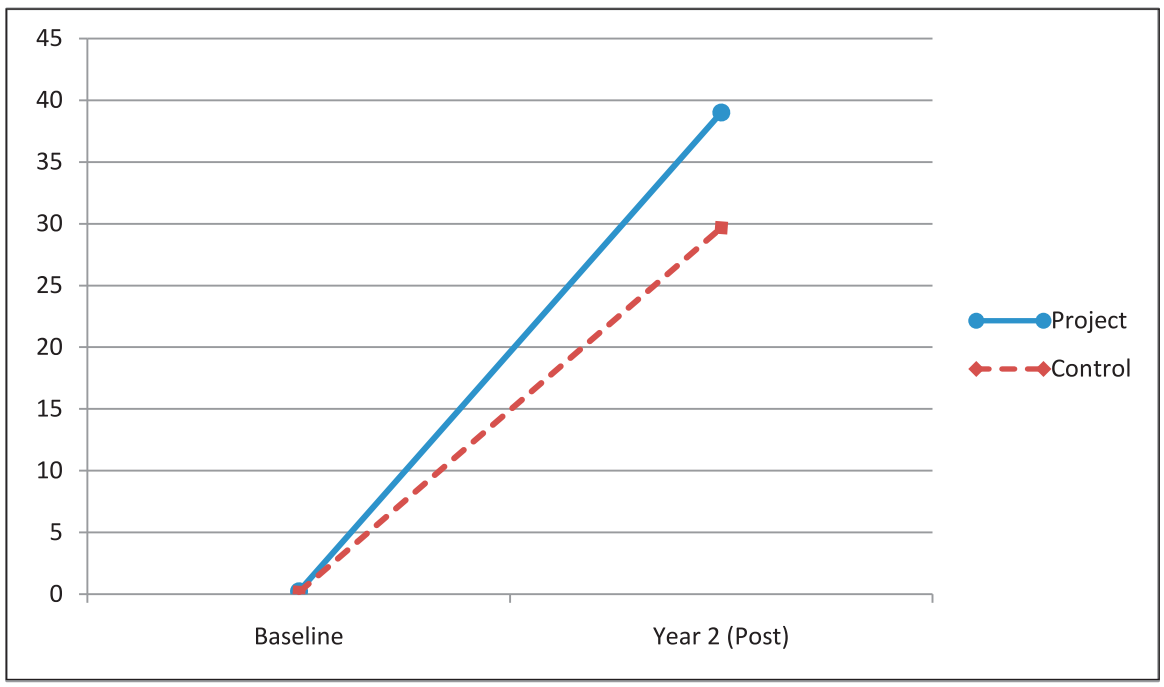

Figure 2: Grade 1 mean learner scores (%s) on the EGRA Word Recognition sub-test over time by school type (for Cohort 2)

25

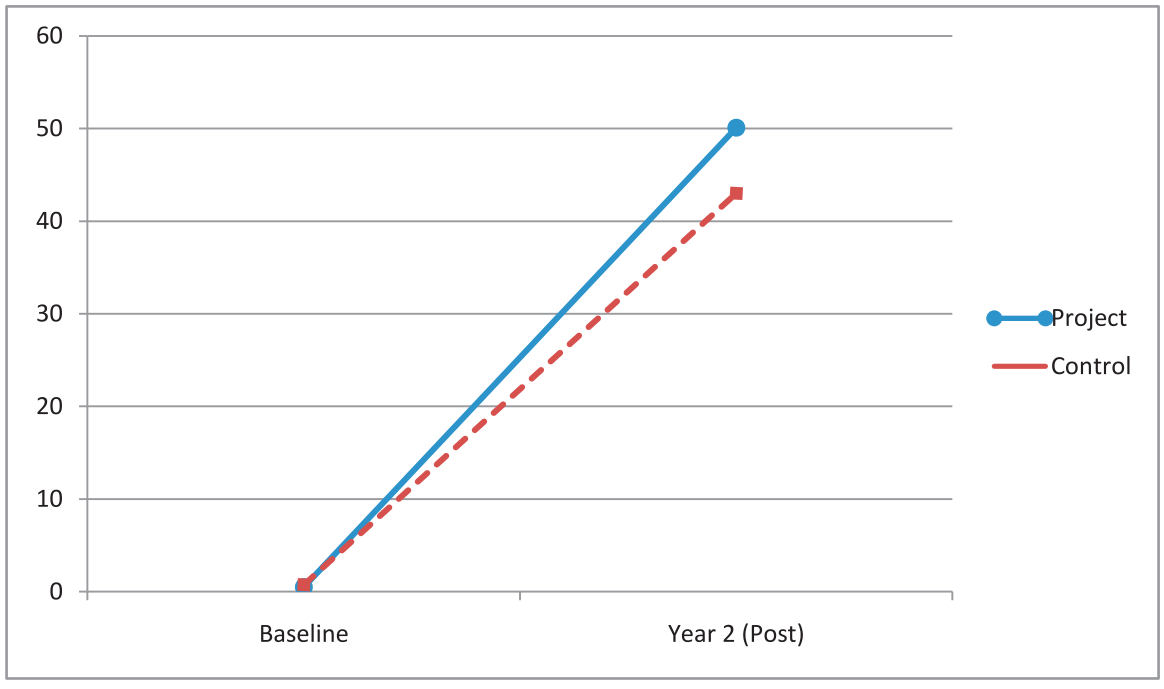

Figure 3: Grade 1 mean learner scores (%s) on the EGRA Non-Word Decoding sub-test over time by school type (for Cohort 2)

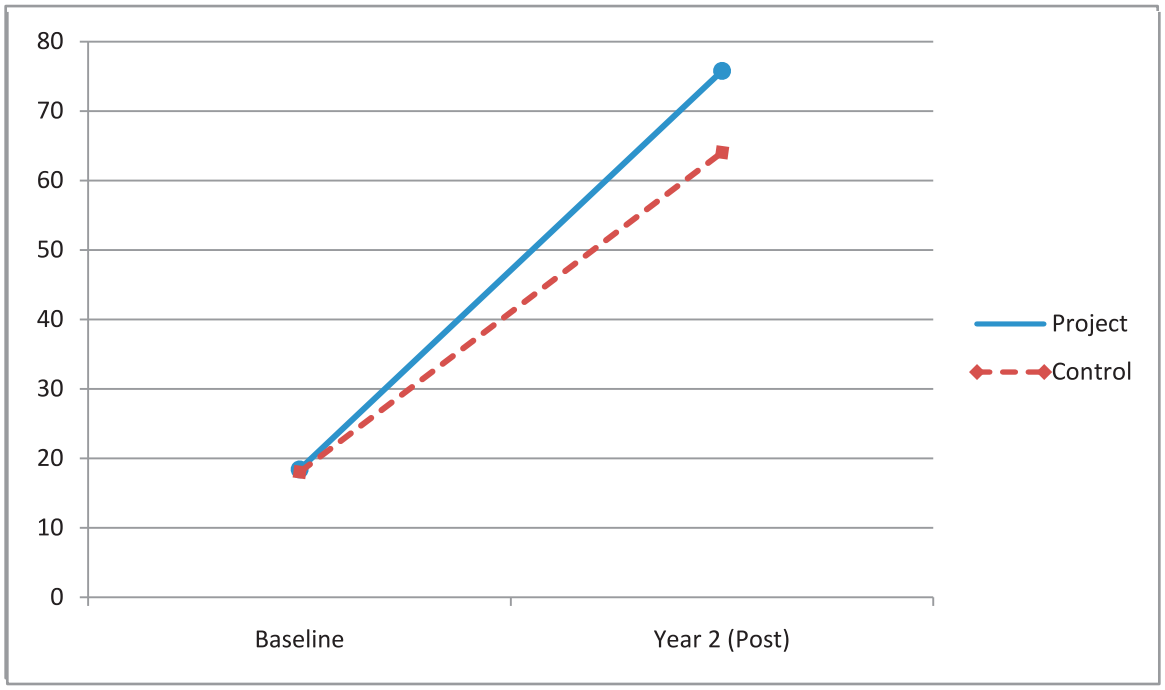

Figure 4: Grade 1 mean learner scores (%s) on the Oral Pictorial sub-test over time by school type (for Cohort 2)

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 6.The results signify that there had been no (statistical) differences between learners at the outset, and that by June differ-

26

ent achievement between children from project and control schools had occurred for the Oral Pictorial test only. However, by November, the table signifies strong gains for project-school learners.

Table 6: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 2014, June 2014 and May 2015

Test (n)	School subset	Baseline	June 2014	May 2015
EGRA Letter Sound (n=688)	Project	6.7	42.7	***76.1
	Control	8.1	34.2	17.5
EGRA Word Recognition (n=689)	Project	0.0	3.7	**36.5
	Control	0.0	2.0	22.8
EGRA Non-WordDecoding (n=684)	Project	0.0	14.4	**56.1
	Control	0.7	15.4	41.6
Oral Pictorial (n=687)	Project	9.3	***67.5	**91.1
	Control	7.4	42.6	82.4

* 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, with the exception of the Oral Pictorial sub-test. 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 45% on these tests, and the fact that fewer than 20% of the learners in two of these three subtests reached an acceptable achievement level by the end of Grade 1. Tables 5 and 6 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 4th 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 652 to 657 for the Schonell Reading Age and EGRA tests and from 639 to 667 for the Literacy test scores. The exact numbers depend on the sub-test. Of these learners, 502 to 507 for the Schonell Reading Age and EGRA tests and 489 to 517 for the Literacy test scores were from project schools. The numbers were 150 for all tests 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.

27

3. Evaluation findings

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. The sets of scores were flagged for possible attrition effects 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. Being in attendance at both testing occasions may have favoured finding higher scores among such learners. However, this effect was highly similar across project- and control- school status. 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 7 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 were greatly in favour of the project group. It is further revealed that statistically significant gain differences occurred over one year for project-school learners in all 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 1 (MCQ), Literacy 2 (sentence ordering) and Literacy 4 (comprehension) sub-tests and the overall literacy score are depicted graphically in Figures 5 to 12 further down. Achievement gains pertaining to the EGRA Letter Sound and Literacy 1 (MCQ) sub-tests were particularly strong. These findings represent improvements ranging from 21% to as high as 262% and over 400% for learners in project schools compared to changes from just more than 10% to almost 190% for learners in the control group. It is noteworthy that the results indicate a negative change of 41.5% on the EGRA Letter Sound sub-test for the control group learners. It is reiterated that Reading Age scores are not measured in percentages, but in years.

Table 7: 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)
Schonell Reading Age (not %)	Project	6.2 (1.4)	***7.5 (1.2)	***1.3 (1.2)	0.6 (0.3)	21.4 (0.9)
	Control	6.2 (1.1)	7.0 (0.7)	0.7 (0.9)		11.6 (0.7)
EGRA Letter Sound	Project	16.5 (17.7)	***84.0 (18.4)	***67.5 (25.7)	74.8 (4.3)	408.7 (3.8)
	Control	17.6 (15.8)	10.3 (8.7)	-7.3 (17.9)		-41.5 (@)
EGRA Word Recognition	Project	58.0 (29.2)	***84.7(20.8)	***26.7 (21.5)	7.8 (0.2)	46.0 (0.9)
	Control	60.1 (26.4)	79.0 (21.9)	18.9 (18.6)		31.5 (0.7)
EGRA Non-Word Decoding	Project	61.0 (24.7)	***85.2 (18.4)	***24.2 (21.5)	14.8 (0.6)	39.7 (1.0)
	Control	62.1 (22.2)	71.5 (17.3)	9.4 (19.8)		15.1 (0.4)
Literacy 1 MCQ	Project	*22.0 (27.1)	***77.0 (30.3)	***55.0 (38.5)	55.0 (1.5)	250.5 (2.0)
	Control	18.2 (24.9)	31.3 (29.0)	13.1 (35.8)		72.0 (0.5)
Literacy 2	Project	14.9 (20.4)	***48.9 (36.4)	***34.0 (40.3)	24.1 (1.1)	227.7 (1.7)
Sentence Order	Control	15.4 (18.9)	25.3 (25.4)	9.9 (27.8)		64.3 (0.5)

28

3. Evaluation findings

Test	School subset	Baseline % (& SD)	End Year 1 % (& SD)	Change % pts (& SD)	Diff-in-diff % pts (& SD)	Change % (& SD)
Literacy 4 Open-ended	**Project**	***9.0 (15.0)	***42.6 (32.4)	***33.7 (30.1)	26.1 (1.3)	375.4 (2.2)
	Control	4.0 (8.5)	11.6 (18.6)	7.6 (19.2)		189.5 (0.9)
Literacy Total	**Project**	**14.3 (19.4)	***51.6 (27.8)	***37.3 (28.3)	28.0 (0.9)	261.7 (1.9)
	Control	9.9 (9.3)	19.2 (17.8)	9.3 (17.2)		94.5 (1.0)

* 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.

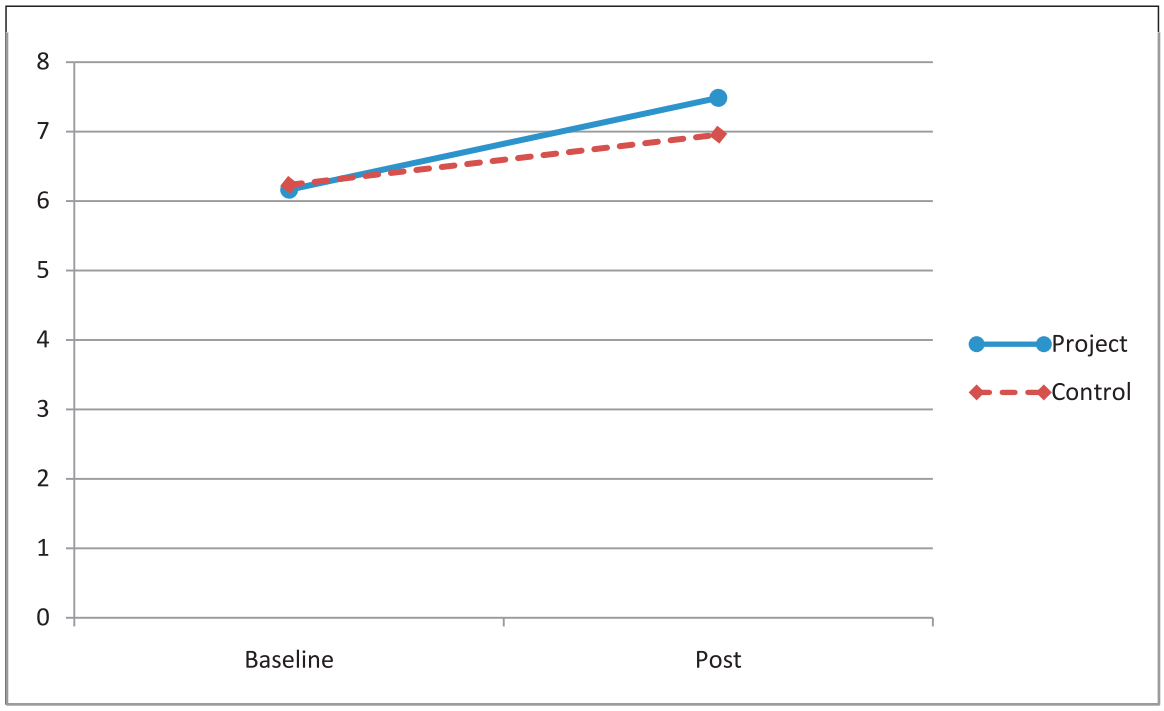

Figure 5: Grade 4 learner mean scores (reading age) on the Schonell Reading Age test over time by school type (for Cohort 2)

29

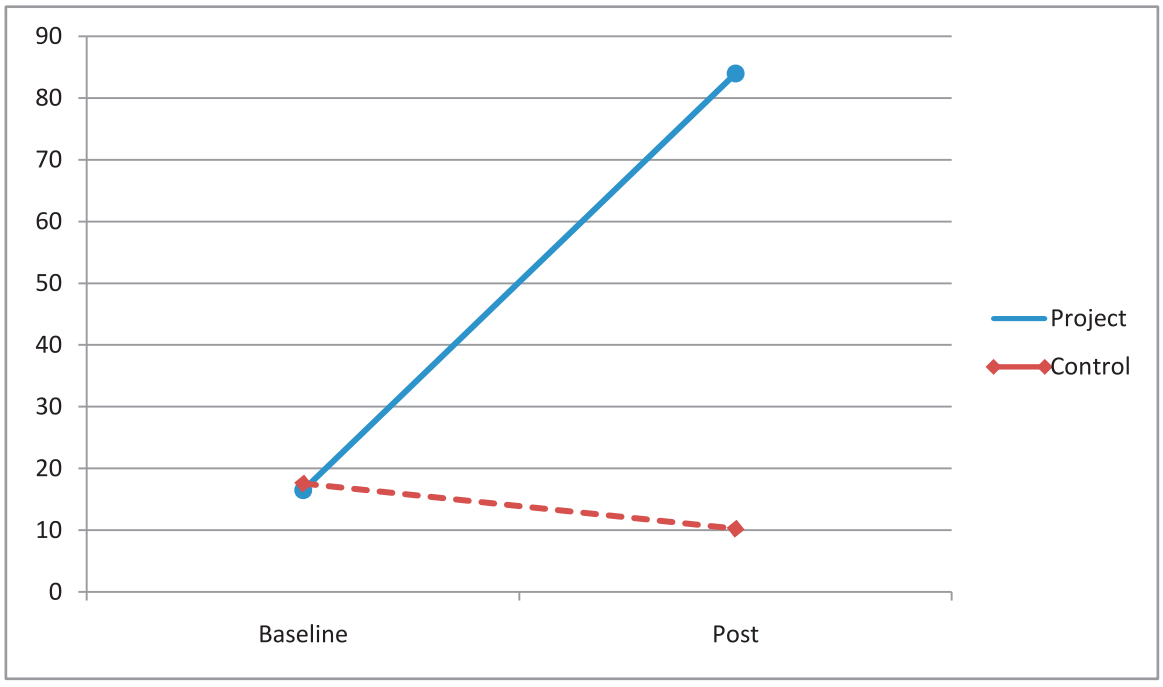

Figure 6: Grade 4 mean learner scores (%s) on the EGRA Letter Sound sub-test over time by school type (for Cohort 2)

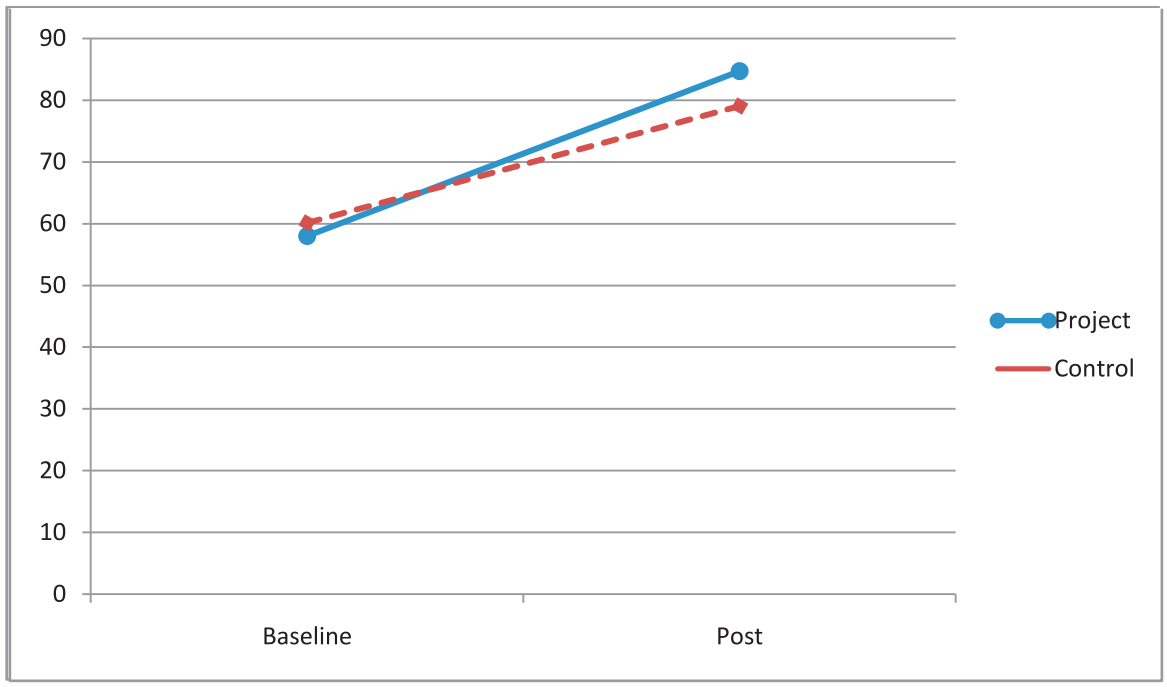

Figure 7: Grade 4 learner mean scores (%s) on the EGRA Word Recognition sub-test over time by school type (for Cohort 2)

30

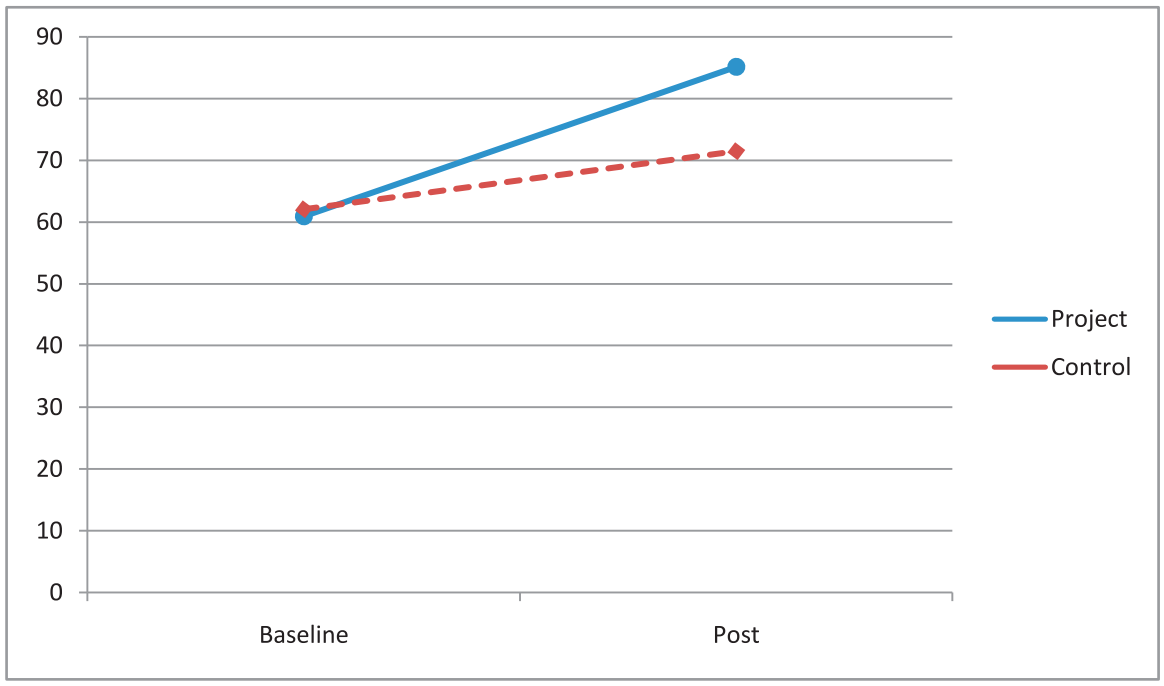

Figure 8: Grade 4 learner mean scores (%s) on the EGRA Non-Word Decoding sub-test over time by school type (for Cohort 2)

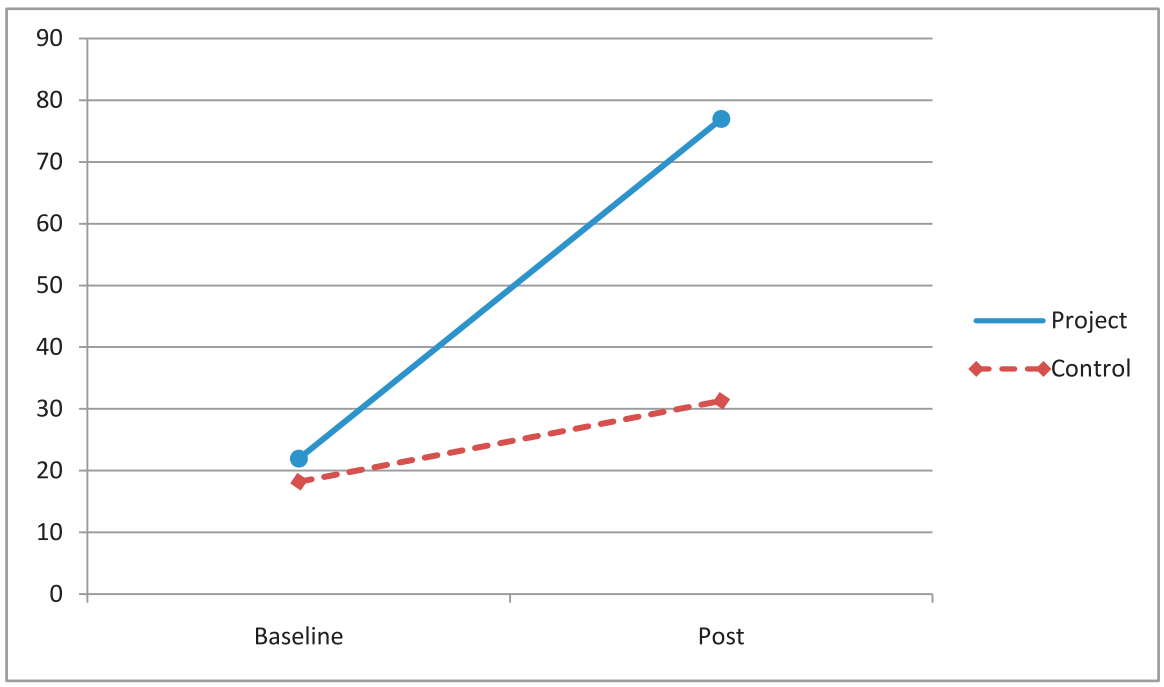

Figure 9: Grade 4 learner mean scores (%s) on the Literacy 1 (MCQ) sub-test over time by school type (for Cohort 2)

31

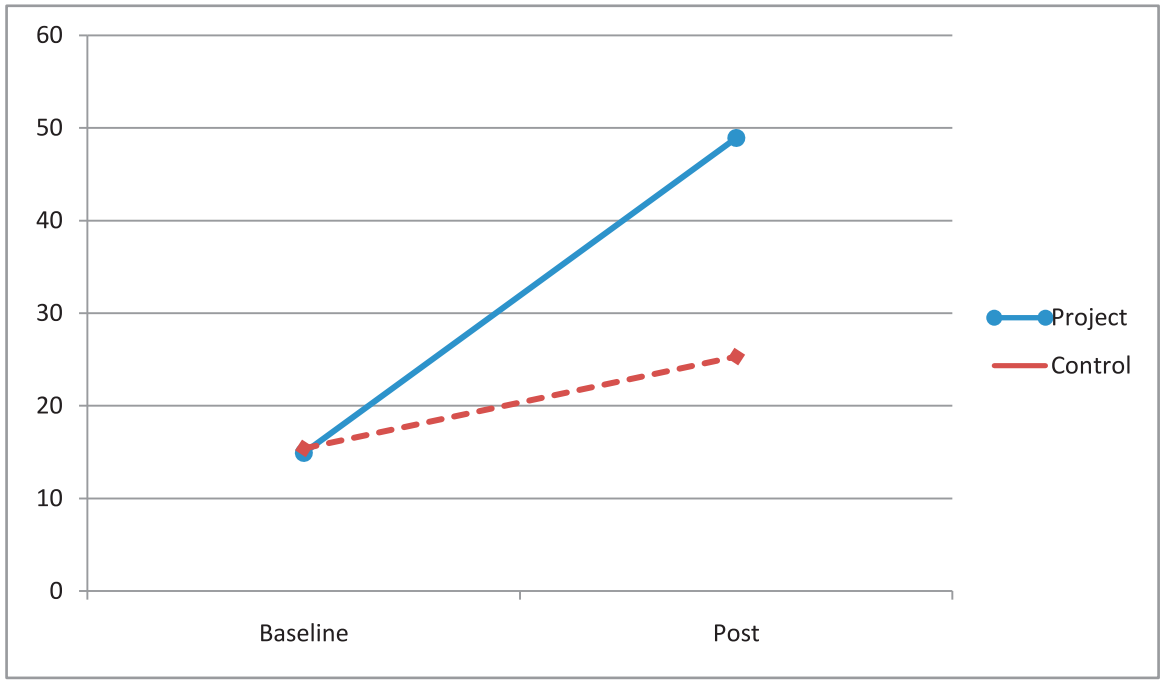

Figure 10: Grade 4 learner mean scores (%s) on the Literacy 2 (Sentence Ordering) sub-test over time by school type (for Cohort 2)

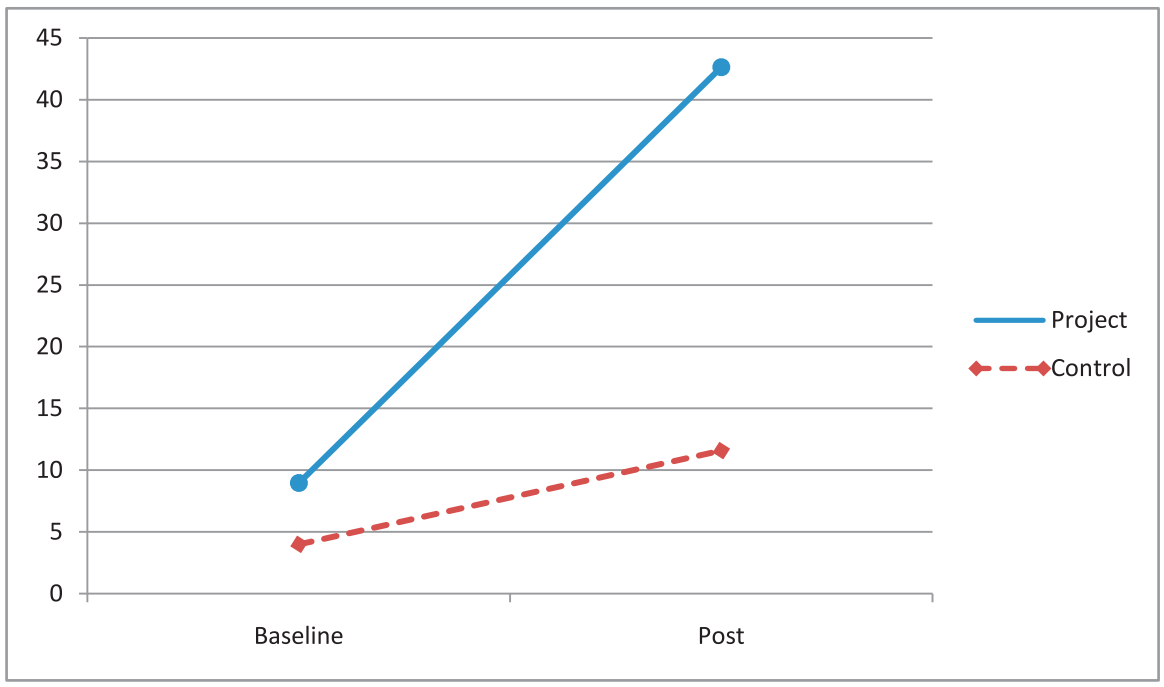

Figure 11: Grade 4 learner mean scores (%s) on the Literacy 4 (Open-ended items) sub-test over time by school type (for Cohort 2)

32

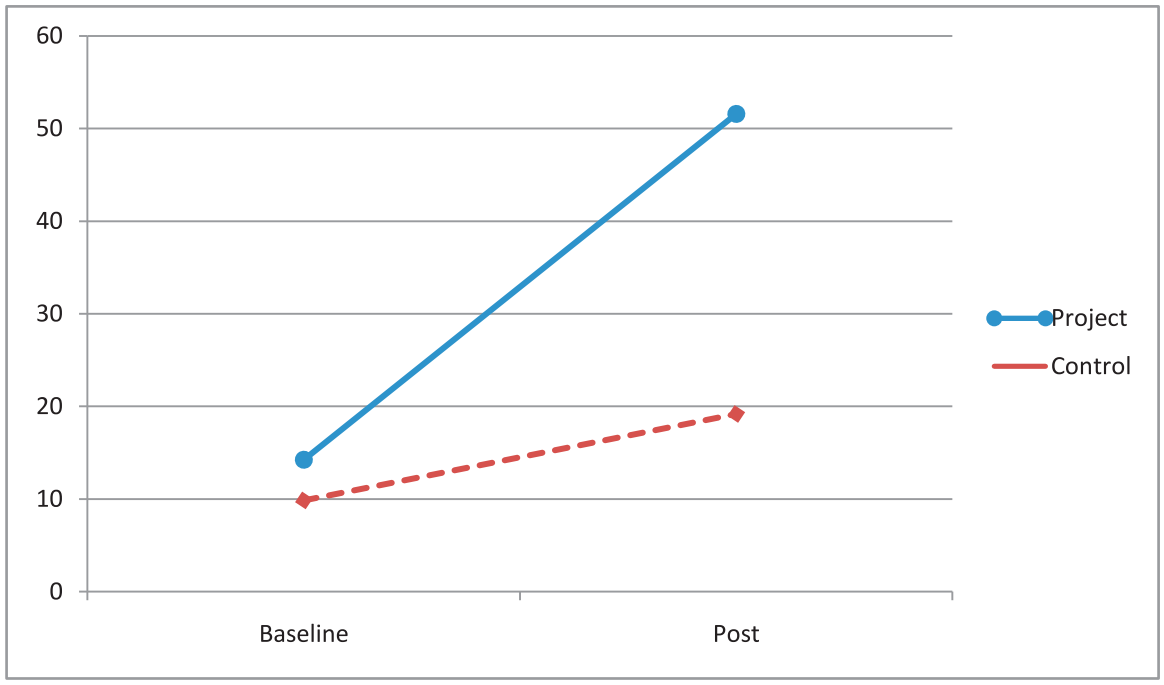

Figure 12: Grade 4 learner mean scores (%s) on the overall (aggregated) Literacy test over time by school type (for Cohort 2)

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 8.

It is evident from Table 8 that after one year a significantly higher proportion of learners in project schools achieved a mark of at least 50% on all tests or a reading age of 8 or higher compared to control-school learners. By mid-year, for the three EGRA sub-tests, learners from project schools showed statistically different higher achievement on the EGRA Letter Sound sub-test than children from control schools. After one year, a much higher achievement outcome was evident for project-school learners in all scores but the EGRA Word Recognition and Non-Word Decoding sub-tests.

Proficiencies within the direct aim of the intervention programme, i.e., phonemic and decoding skills, had either a low score at baseline, i.e. the EGRA Letter Sound sub-test which continued to increase or a high score at baseline, i.e. the EGRA Word Recognition and Non-Word Decoding sub-tests, and then increased during the course of the year. However, as expected, proficiencies in derived competencies, such as comprehension, increased after a time lag.

Table 8: 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 2014, June 2014 and May 2015

Test (n)	School subset	Baseline	June	May
Schonell Reading Age of 8 yrs (n=652)	Project	0.0	**7.2	***32.1
	Control	0.0	1.3	4.7
EGRA Letter Sound (n=654)	Project	9.0	***52.2	***95.0
	Control	6.0	2.0	0.0
EGRA Word Recognition (n=657)	Project	64.1	73.6	92.0
	Control	69.3	68.7	90.7
EGRA Non-Word Decoding (n=654)	Project	74.8	79.2	95.0
	Control	77.3	72.7	92.0
Literacy 1 MCQ (n=647)	Project	16.3	***26.2	***79.7
	Control	12.7	9.3	25.3

33

Test (n)	School subset	Baseline	June	May
Literacy 2 Sentence Ordering (n=639)	Project	14.7	21.5	***54.4
	Control	13.3	20.0	29.3
Literacy 4 Open-ended (n=665)	Project	*4.1	**7.6	***42.1
	Control	0.7	0.7	5.3
Literacy Total (n=667)	Project	*4.1	***8.5	***50.1
	Control	0.7	0.7	6.7

* 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 0,7 of a year over the course of a year. In effect those learners therefore fell behind slightly further and are now just over a year, at minimum, behind where they should have been. In contrast, learners from project schools caught up more than a year, and are now just over half a year behind where they should have been. Tables 7 and 8 provide the necessary details underpinning this observation.

3.2.3 Grade 7

The total sample of Grade 7 learners was taken from all schools included in the intervention, thereby encompassing schools from the first cohort in 2013, 11 project and five control schools, and from the second cohort in 2014, 16 project and four control schools. Once Grade 7 learners have completed their school year they transfer to high school, effectively ending their participation in the intervention. Therefore, Grade 7 learners from the first cohort schools and Grade 7 learners from the second cohort schools were in the intervention for the same period of one year during 2014. During analysis of the results, all Grade 7 learners from 36 schools were treated as one group who completed a year of the intervention during 2014. Note that a few of these schools are Junior Primary Schools without Grade 7 learners.

Seven tests were again 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 4th 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 1 572 learners for the Schonell Reading Age assessment and 1 583 learners for the various Literacy scores. The exact numbers depend on the sub-test. Of these learners, 1 100 to 1 103 for the Schonell Reading Age and EGRA tests and 1 078 to 1 102 for the Literacy test scores were from project schools. The numbers were either 470 or 471 for the Schonell Reading Age and EGRA tests and from 450 to 481 for the Literacy test scores for control-school learners.

34

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. Very few consistent and strong differences were observed, in relation to comparing paired and not-paired sets of data within and across the points of testing and control and project status. No attrition bias could therefore be detected. 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 9. 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 the baseline scores favoured learners from the project in both the EGRA Letter Sound sub-test and Literacy 4 Open-ended sub-test. By November, all statistical differences favoured project-school learners. The gain scores further confirm the statistically significant greater intervention benefits to project-school learners on all eight test scores over the period of one year. The control-school learners have been significantly overtaken by the project group in all tests. 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 13 to 20.These findings represent improvements ranging mostly between 10% and 65% for learners in project schools compared to changes from 3% to 30% for learners in the control group.

Table 9: 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)
Schonell Reading Age (not%)	Project	7.3(1.1)	***8.0 (1.3)	***0.7 (1.0)	0.5 (0.5)	9.6 (0.6)
	Control	7.5 (3.5)	7.7 (1.1)	0.2 (3.4)		2.7 (0.1)
EGRA Letter Sound	Project	***13.9 (21.6)	***85.9 (17.2)	***72.0 (26.5)	56.5 (2.3)	518.0 (3.3)
	Control	9.9 (15.0)	25.4 (23.2)	15.5 (24.1)		156.6 (1.0)
EGRA Word Recognition	Project	81.3 (22.1)	*90.0 (17.2)	***8.6 (13.2)	3.0 (0.1)	10.6 (0.4)
	Control	82.1 (19.5)	87.8 (16.5)	5.6 (13.0)		6.8 (0.3)
EGRA Non-Word Decoding	Project	72.6 (21.5)	***82.8 (18.5)	***10.1 (17.2)	5.1 (0.2)	13.9 (0.5)
	Control	73.2 (19.4)	78.2 (17.6)	5.0 (18.5)		6.8 (0.3)
Literacy 1 MCQ	Project	40.0 (31.0)	***66.0 (26.9)	***26.0 (34.6)	13.3 (0.4)	65.0 (0.8)
	Control	41.1 (30.1)	53.8 (31.2)	12.7 (35.0)		30.9 (0.4)
Literacy 3 Write Paragraph	Project	21.3 (25.6)	***33.7 (29.8)	***11.9 (32.0)	12.0 (0.5)	55.9 (0.5)
	Control	20.7 (24.2)	19.9 (23.6)	-0.1 (25.8)		-0.5 (0.0)
Literacy 4 Open-ended	Project	*13.7 (16.7)	***18.2 (17.8)	***4.5 (19.7)	4.9 (0.3)	32.8 (0.3)
	Control	11.6 (16.3)	11.2 (14.8)	-0.4 (16.2)		-3.4 (@)
Literacy Total	Project	23.3 (15.9)	***35.7 (17.6)	***12.3 (15.6)	8.7 (0.6)	52.8 (0.8)
	Control	22.1 (16.2)	25.7 (17.1)	3.6 (14.3)		16.3 (0.2)

* 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 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.

35

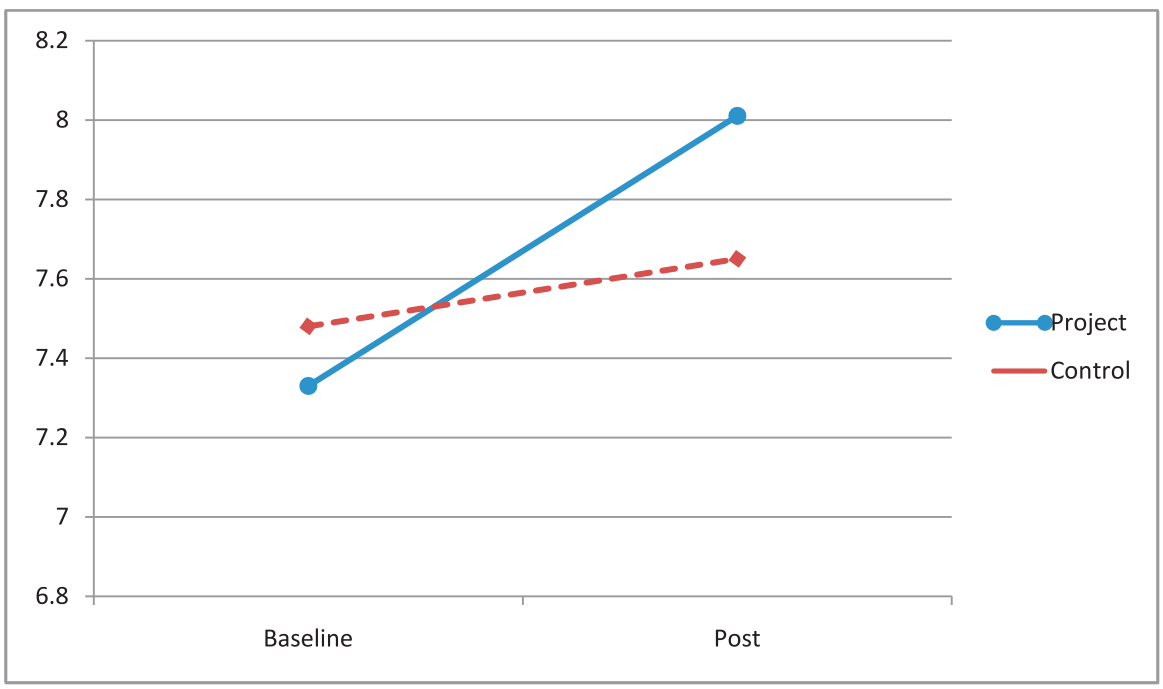

Figure 13: Grade 7 mean learner scores (age) on the Schonell Reading Age test over time by school type (for Cohort 2)

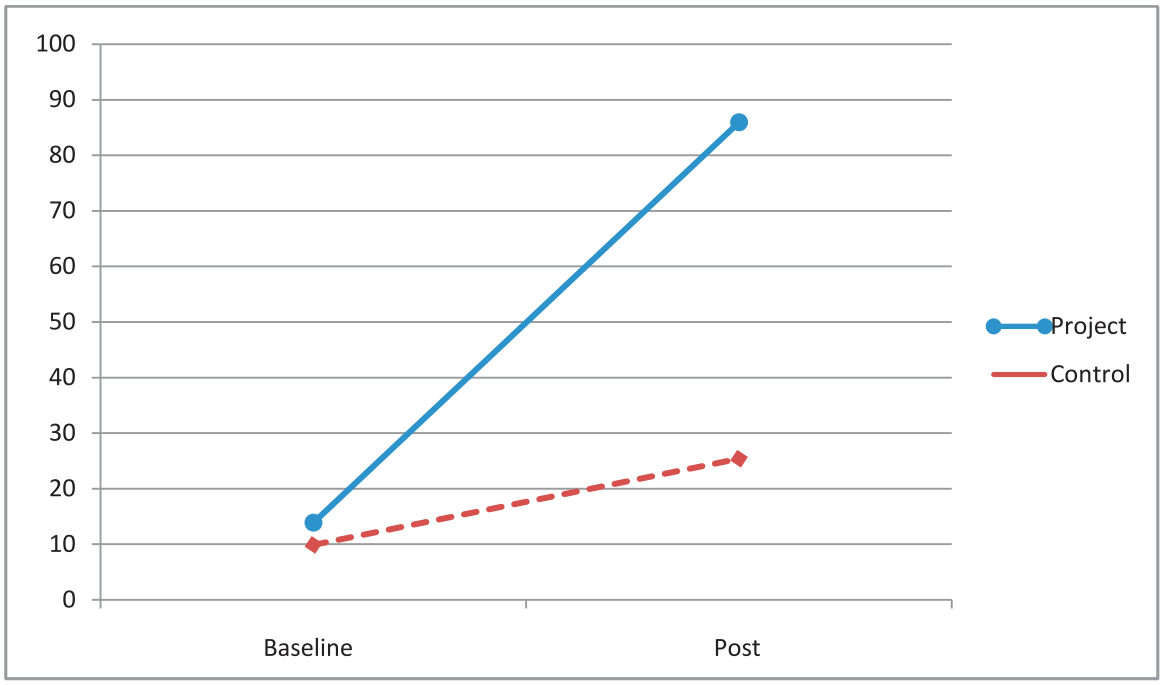

Figure 14: Grade 7 mean learner scores (%s) on the EGRA Letter Sound sub-test over time by school type (for Cohort 2)

36

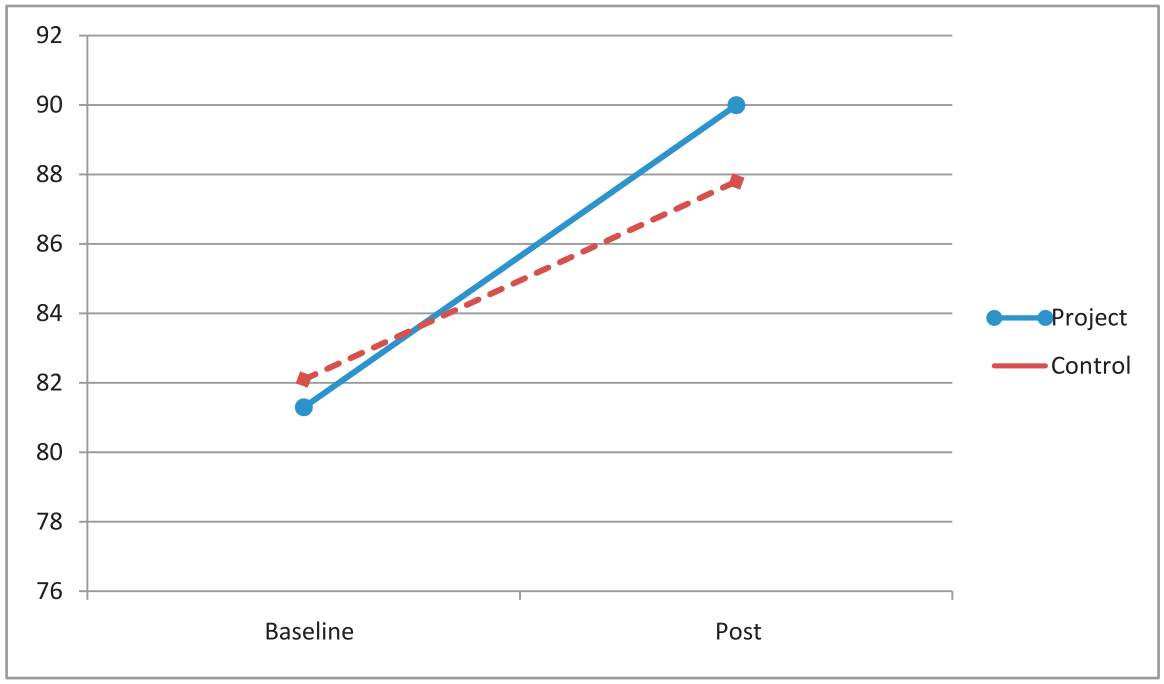

Figure 15: Grade 7 learner mean scores (%s) on the EGRA Word Recognition sub-test over time by school type (for Cohort 2)

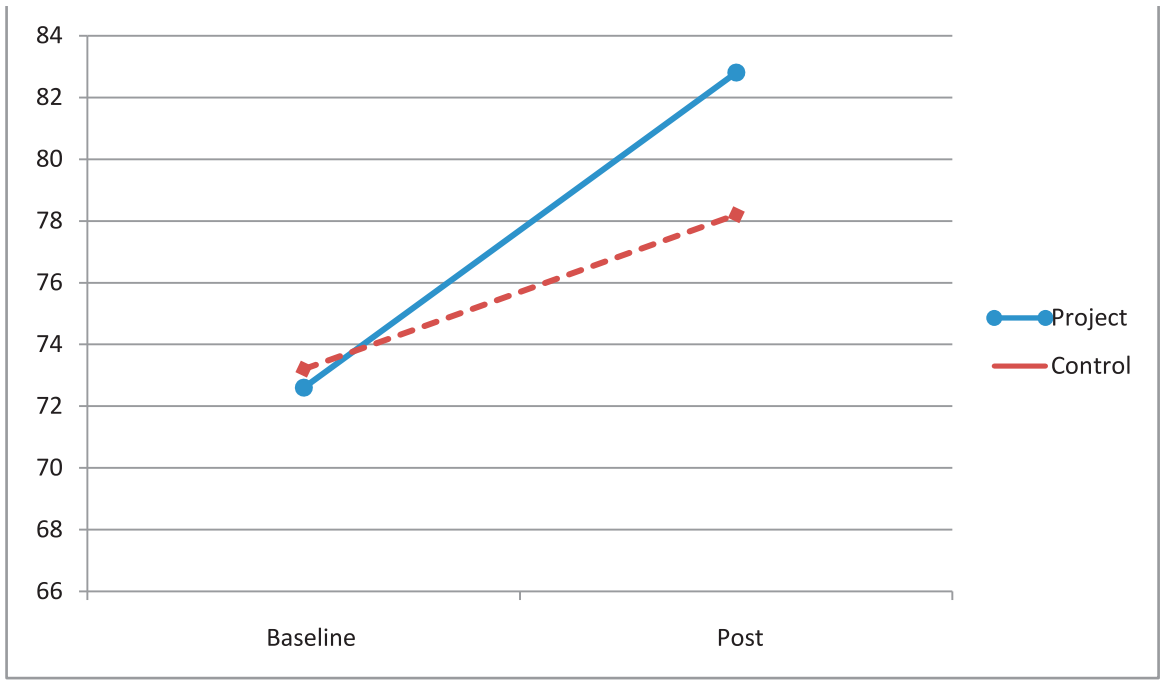

Figure 16: Grade 7 learner mean scores (%s) on the EGRA Non-Word Decoding sub-test over time by school type (for Cohort 2)

37

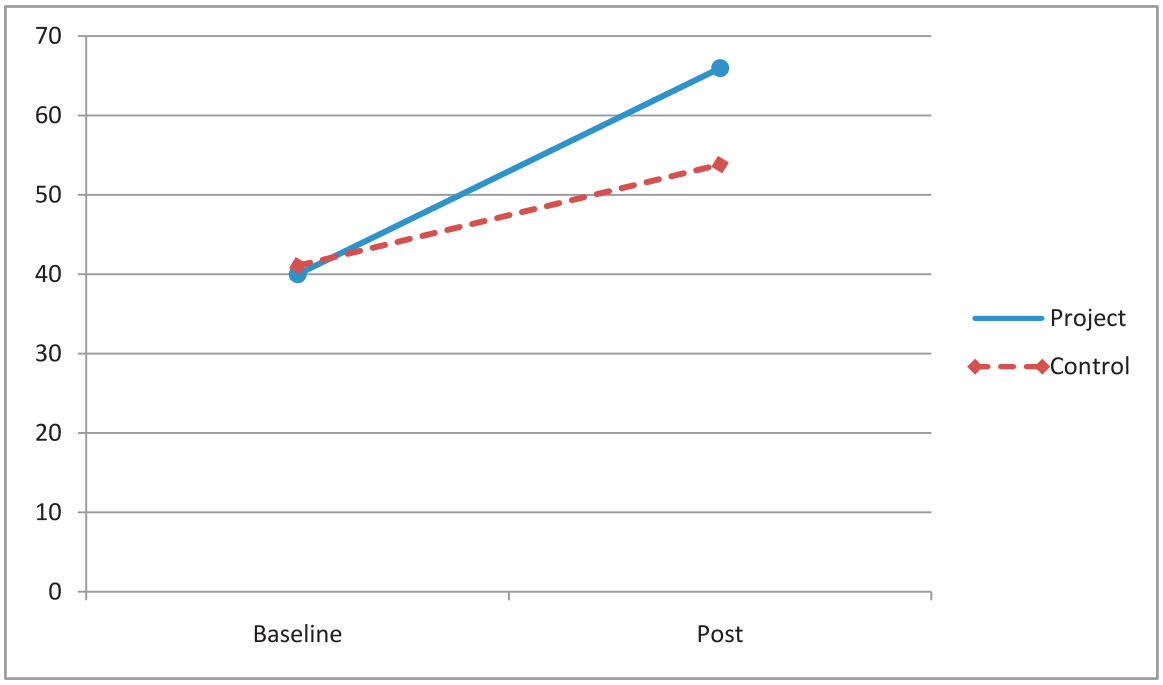

Figure 17: Grade 7 mean learner scores (%) on the Literacy 1 sub-test (MCQ) over time by school type (for Cohort 2)

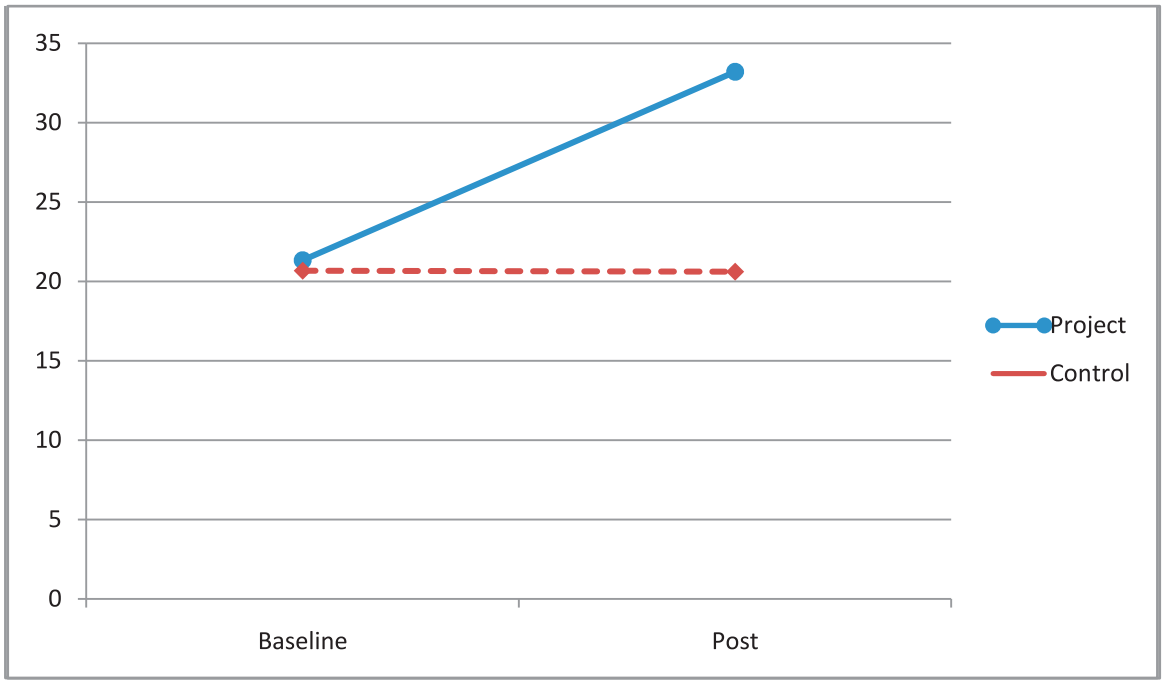

Figure 18: Grade 7 mean learner scores (%) on the Literacy 3 sub-test (Paragraph Writing) over time by school type (for Cohort 2)

38

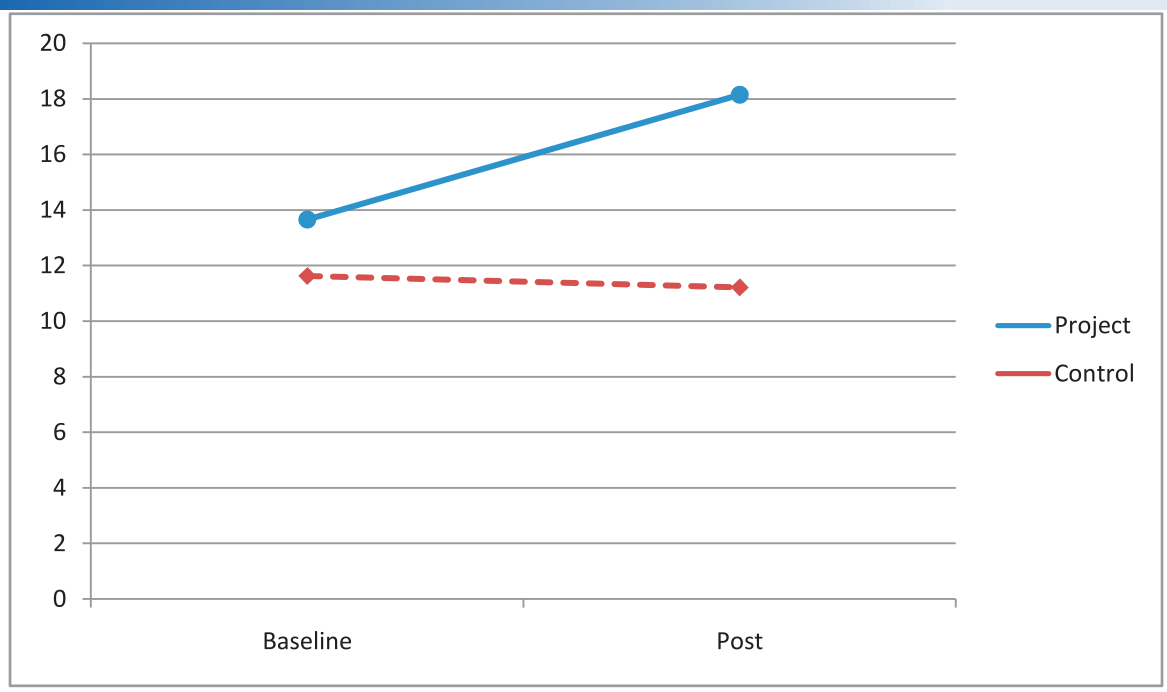

Figure 19: Grade 7 mean learner scores (%) on the Literacy 4 sub-test (Open-ended items) over time by school type (for Cohort 2)

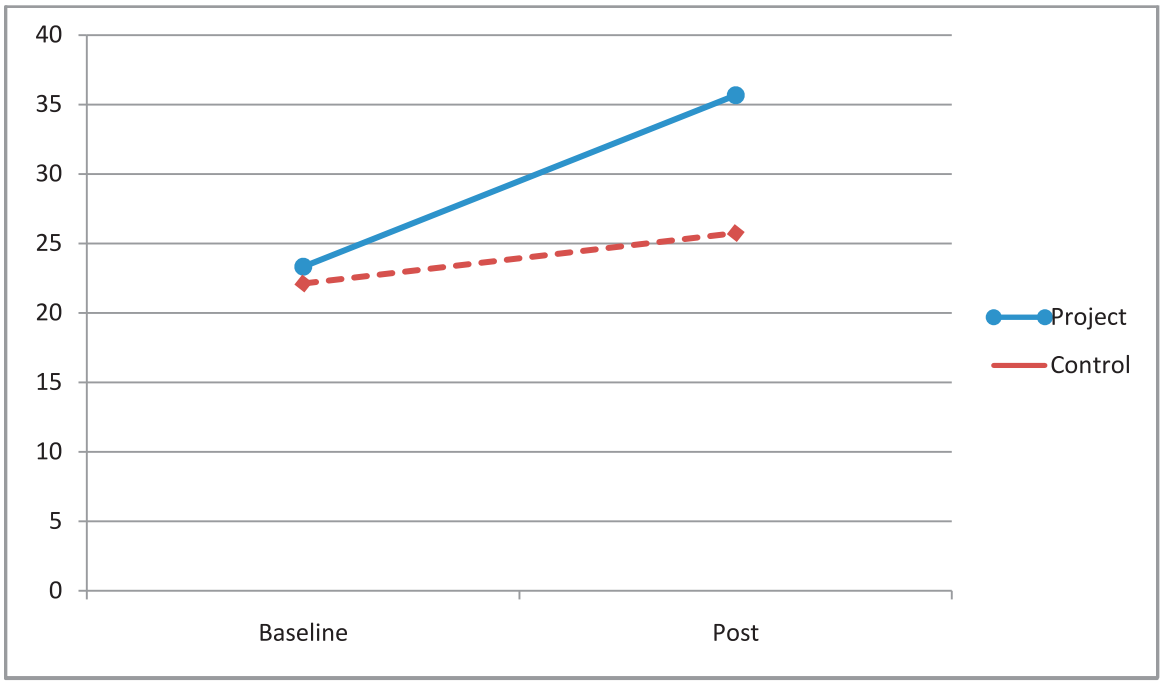

Figure 20: Grade 7 mean learner scores (%) on the total Literacy test over time by school type (for Cohort 2)

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 10 reflects the situation. The baseline situation in February favoured the project groups, but only at the 95% probability level. At the mid-year assessment in June the learners in project schools significantly surpassed learners in control schools. This situation was again reflected when by November, after one year of intervention, significantly more project-school learners were achieving at least at the 50%-mark with the exception of the EGRA Word Recognition and Non-Word Decoding sub-tests as well as Literacy 1 MCQ

39

and Literacy 4 Open-ended items sub-tests where there was no significant difference between project-school and control-school learners.

Table 10: 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 2014

Test (n)	School subset	Baseline	June	November
Schonell Reading Age of 11 yrs (n=1572)	Project	*1.0	*1.0	***3.0
	Control	0.0	1.0	2.0
EGRA Letter Sound((n=1571)	Project	*10.0	***84.0	***96.0
	Control	6.0	13.0	20.0
EGRA Word Recognition (n=1573)	Project	90.0	*90.0	96.0
	Control	92.0	86.0	96.0
EGRA Non-Word Decoding (n=1574)	Project	88.0	**89.0	94.0
	Control	89.0	84.0	93.0
Literacy 1 MCQ (n=1583)	Project	36.5	**57.4	100.0
	Control	37.8	45.1	100.0
Literacy 3 Paragraph writing (n=1583)	Project	24.1	***27.6	***40.2
	Control	21.8	17.1	16.4
Literacy 4 Open-ended items (n=1583)	Project	4.6	10.1	7.1
	Control	4.4	6.0	3.5
Literacy Total (n=1583)	Project	8.1	***18.1	***22.2
	Control	7.3	9.6	10.6

* 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

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 less than a quarter of 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 0,7 years, which is less than the required minimum of one year. They are therefore still at least three years behind the required level. Put differently, after a full seven or eight years of schooling, only 2-3% 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 2014 made much stronger inroads into their reading backlogs. In addition, the intervention programme seems to have succeeded in starting to address some crucial elements of learner comprehension and writing ability. Tables 9 and 10 provide the necessary details underpinning these comments.

The Grade 7 learner group was comprised of learners from schools which had taken part in the first year of the intervention (2013) as well as new schools recruited for expansion of the intervention in the second year (2014). When Grade 6 learners from the first year schools progressed to Grade 7 they became active participants in the intervention. However, as they were already enrolled in schools taking part in the intervention when they were in Grade 6, they may have been inadvertently influenced by the intervention before their official participation. Therefore, the Grade 7 learner achievements were compared between learners from the first and second year schools. The resulting mean percentage scores are reflected in Table 11.

40

Table 11: Mean percentage achievement (age estimate for Schonell test) during testing in February and November 2014 as well as percentage increase for Grade 7 learners by school type (2013/2014)

Test (n)	School subset	First intervention year	Baseline	% increase Feb-Nov 2014	November
Schonell Reading Age of 11 yrs (n=1572)	Project	2013 (Coh 1)	***7.5 (1.1)	0.5 (0.9)	8.0 (1.2)
		2014 (Coh 2)	7.2 (1.2)	***0.8 (1.0)	8.1 (1.4)
	Control	2013 (Coh 1)	7.6 (4.3)	0.1 (4.2)	7.7 (1.1)
		2014 (Coh 2)	7.3 (0.8)	0.3 (0.6)	7.6 (0.9)
EGRA Letter Sound (n=1571)	Project	2013 (Coh 1)	***21.4 (28.2)	63.4 (30.7)	84.8 (17.3)
		2014 (Coh 2)	7.9 (11.2)	***78.7 (20.3)	86.6 (17.1)
	Control	2013 (Coh 1)	8.6 (14.6)	**17.9 (24.8)	26.5 (23.8)
		2014 (Coh 2)	*12.3 (15.4)	10.8 (22.0)	23.0 (22.1)
EGRA Word Recognition (n=1573)	Project	2013 (Coh 1)	82.5 (21.0)	7.3 (13.1)	89.8 (17.0)
		2014 (Coh 2)	80.5 (22.9)	**9.6 (13.2)	90.1 (17.4)
	Control	2013 (Coh 1)	81.2 (20.8)	6.1 (13.0)	87.3 (17.6)
		2014 (Coh 2)	83.5 (17.5)	4.5 (12.9)	88.1 (15.8)
EGRA Non-Word Decoding (n=1574)	Project	2013 (Coh 1)	**74.9 (20.9)	8.0 (18.0)	82.9 (17.8)
		2014 (Coh 2)	71.1 (21.7)	***11.8 (17.3)	82.8 (18.9)
	Control	2013 (Coh 1)	72.0 (20.4)	5.5 (19.2)	77.5 (18.5)
		2014 (Coh 2)	75.1 (17.9)	4.1 (17.2)	79.2 (16.0)
Literacy 1 MCQ (n=1583)	Project	2013 (Coh 1)	**43.2 (30.4)	20.7 (35.4)	63.9 (27.2)
		2014 (Coh 2)	37.5 (31.2)	***30.1 (33.4)	*67.6 (26.6)
	Control	2013 (Coh 1)	*43.4 (30.3)	10.7 (35.3)	54.1 (31.0)
		2014 (Coh 2)	36.5 (29.3)	16.8 (34.1)	53.3 (31.8)
Literacy 3 Paragraph writing (n=1583)	Project	2013 (Coh 1)	19.4 (24.4)	12.7 (31.5)	32.1 (28.9)
		2014 (Coh 2)	*22.9 (26.5)	11.2 (32.5)	34.1 (29.8)
	Control	2013 (Coh 1)	21.0 (25.4)	0.1 (25.2)	21.1 (25.3)
		2014 (Coh 2)	19.9 (21.2)	-0.5 (27.3)	19.4 (20.2)
Literacy 4 Open-ended items (n=1583)	Project	2013 (Coh 1)	11.6 (14.2)	*5.9 (18.5)	17.5 (16.1)
		2014 (Coh 2)	***15.3 (18.2)	3.4 (20.6)	18.6 (18.8)
	Control	2013 (Coh 1)	10.9 (15.6)	*0.8 (16.1)	11.8 (15.7)
		2014 (Coh 2)	13.0 (17.5)	-2.9 (16.1)	10.1 (12.9)
Literacy Total (n=1583)	Project	2013 (Coh 1)	23.0 (15.0)	11.4 (16.1)	34.5 (16.7)
		2014 (Coh 2)	23.6 (16.6)	13.0 (15.2)	*36.6 (18.3)
	Control	2013 (Coh 1)	22.8 (16.6)	3.6 (15.0)	26.4 (18.0)
		2014 (Coh 2)	20.8 (15.2)	3.8 (13.0)	24.5 (15.3)

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

For the control schools, there were mostly marginal differences between the schools from the first year and second years of the intervention. For the few significant differences, these were most likely random. However, the results indicate that there was an effect on student achievement where learners from project-schools which had undergone the first year of intervention started off from higher baseline scores (statistically significant) in the Schonell Reading Age test, the EGRA Letter Sound and Non-Word Decoding sub-tests, and the Literacy 1 sub-test. The second year project-school learners achieved higher scores on the Literacy 3 and Literacy 4 sub-tests which indicate that the contextual effects of the intervention were limited to proficiencies within the direct aim of the intervention programme, i.e. phonemic and decoding skills, rather than the derived competencies, such as comprehension. When assessed in November 2014, there were no statistically significant differences in six of the eight sub-tests in project-school learners' achievement when compared by intervention year, including sub-tests where project-school learners in the first year of the intervention had achieved higher baseline scores. In the Literacy 1 sub-test and the Literacy Total score, project-school learners from the second year of the intervention actually achieved higher (*p<05) scores than their first year counterparts. The results suggest that there was some crossover from the intervention

41

programme to the Grade 6 learners in the first year project-schools. However, the percentage increase for these learners was eclipsed by that of learners in the second year project-schools where November 2014 scores were mostly equivalent. The intervention programme therefore improved learner scores to the required level independent of the baseline score with which the learner entered the programme.

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

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 27 January to 18 February 2014 (i.e., the baseline assessment), from 10 to 20 June 2014 (i.e., mid-term or Semester 1 assessment), from 5 to 14 May 2015 for the Grade 1 and 4 learners (end-of-Year 2 assessment).

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 the 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 2014, June 2014, and May 2015, some schools on sJsK's normal assessment schedule were visited by a team of HSRC staff. As far as possible, all test contents were covered in the process for a grade / class group. 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 12. As reflected in Table 12, the Grade 7 schools were not monitored during November 2014. As mentioned previously, the Grade 7 learners were abruptly assessed in November 2014 and HSRC staff was unable to organise monitoring in time.

Table 12: Monitoring coverage at the baseline and after the first semester and first year

		Baseline: February 2014		Midline: June 2014		Endline: May 2015	
		% monitored		% monitored		% monitored	
School type	Gr.	Overall	Per school type	Overall	Per school type	Overall	Per school type
Project	1	(15.6)	15.8	(38.6)	29.8	(27.8)	16.0
Control			15.0		70.9		70.9
Project	4	(20.1)	14.0	(27.5)	16.3	(28.3)	17.2
Control			43.6		70.5		70.5

42

		Baseline: February 2014		Midline: June 2014		Endline: May 2015	
		% monitored		% monitored		% monitored	
Project	7	(7.6)	7.0	(10.9)	6.6	-	-
Control			9.3		22.8		-

Specific details regarding these forms and quality assurance of the test-administration and scoring procedures are provided in the Year 1 report.

Brief notes on the investigation of assessment bias in relation to learners from project and control schools follow, and the summaries in Annexures 4 to 6, which reflect the potential impact of monitoring status at baseline (February 2014), and the May 2015 end-line, or neither, either or both occasions, on the outcomes in terms of learner achievement scores and gains over time. The by now familiar caveat is repeated that over-reliance on the statistical indications provided in the annexures is not advisable given the fact that corrections for the clustered design feature of the study would render by far a majority of the "significant" indications otherwise. One should rather note the trends at play. These can be identified, as before, in that 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 **). Below follow brief indications of key observations by grade. Only salient observations and overall trends are highlighted given the vast number of calculations and figures. Complete analyses were nevertheless conducted systematically and scrutinised in order not to overlook anything important. The focus was more on learner gain scores than on specific learner score levels in February 2014 and May 2015.

3.3.1 At Grade 1 level

There were several possibly meaningful (significant) findings yielded in relation to learner achievement gain scores over time between monitored and not-monitored schools for project-school and control-school learners. It appeared as though there could have been bias in the control-school learner scores. Although indications of possible scoring bias could exist in relation to assessment and scoring of learners from the control schools, these effects are only partially confirmed when looking at project-school figures, where counterintuitive findings for the same tests were observed. The sole indication of bias in the project-schools, regarding the gain and end-of-year scores for the Oral Pictorial test in the February monitoring period, is also unconfirmed as there was no indication of bias for the same in the control-school scores.

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

- All except one of the observed data trends for gain scores and end of year scores significantly different between monitored and not monitored are counterintuitive; they indicate opposite trends to what would constitute assessment or scoring bias attempting to show greater intervention effects.

For achievement scores and gains among learners in the control group

- As noted above, the observed differences which could indicate bias in the gain and end of year scores are not supported by the corresponding project-school scores.

43

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, Annexures 5a to 5b largely display what can be considered a random pattern. It is characterised by a majority of counterintuitive observations, as well as inconsistent observations across project and control groups, and over testing time. Main observations on the basis of Annexures 5a and 5b for the full-year monitoring data follow.

For achievement scores and gains among learners in the project group

- Particularly for Schonell Reading Age test and EGRA sub-tests, observed achievement score patterns are quite inconsistent between monitored and not-monitored learners across assessment points (February 2014 and May 2015).
- Most especially for the Literacy sub-tests, observed achievement score patterns for each test are inconsistent where the baseline score indicates bias but the gain and end of year scores are counterintuitive.

For achievement scores and gains among learners in the control group

- There were very few meaningful (significant) differences observed in the control-school learner data.
- Several of the differences on the basis of monitoring status are counterintuitive, especially for February 2014 scores and gain scores over time, and in particular for the Schonell Reading Age test and EGRA sub-test scores.
- Score patterns are contradictory and inconsistent across assessment points (February 2014 and May 2015).
- Data trends are inconsistent or contradictory across sub-tests.

3.3.3 At Grade 7 level

There were very few meaningful (significant) differences in relation to learner achievement gain scores over time as compared between monitored and not-monitored schools across project and control learners. In addition, Annexures 6a to 6b again largely display what can be considered a random pattern. It is characterised by a majority of counterintuitive observations, as well as inconsistent observations across project and control groups, and over testing time. Main observations on the basis of Annexures 6a and 6b for the full-year monitoring data follow. It should be noted that there are no results for the influence of monitoring on either or both occasion monitoring dates. The final monitoring could not be performed due to factors which have been previously explained.

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.

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.

44

3.3.4 Concluding remarks

It is concluded that the observed patterns and effects across grades, (sub-)tests, assessment points, monitoring points and project status were too erratic, inconsistent and contradictory to indicate assessment or scoring bias. In addition, the high number of counterintuitive observations supports the interpretation that there were no deliberate attempts during assessment or scoring with the aim of making the impact of the intervention appear more positive. The observed patterns, in a quantitative sense, are likely to be random. Therefore, no conscious, concerted and consistent effort could be observed to indicate that initial scores were manipulated at the project-schools to be especially low when assessment was not monitored at intervention schools, or likewise especially high at final assessment, and as a result also as gain scores over time. Neither did the inverse become apparent in relation to the unmonitored control-school learners. The evidence at hand can be taken to indicate that assessment and scoring were not biased, because being monitored has not changed any patterns substantively, consistently or purposively.

There are two further observations or arguments which support the conclusion above. First, many scales are quite short, i.e. have only a few items. This implies that small point-allocation differences result in fairly large percentage-scale changes. Larger fluctuations may then be the result. One could easily attribute these to deliberate scoring bias, although they could possibly still be largely random scale artefacts. Second, monitoring may remove deliberate or unconscious scoring bias, but also increase concentration and scoring quality in others. These effects could be in opposite directions, thus increasing inconsistencies. To explain further, scorers could miss either wrong or correct answers when not being monitored. Being monitored may reduce either. These two points comprise further caution against assuming 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. Without presenting detailed tables, core statistics appear in the endnotes.

4. Discussion of results

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 serves to inform similar new programmes 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 affirmative, 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 for the second cohort, as for the first, only when data were available from both the baseline (February 2014) and the end-line (November 2014 or May 2015) assessments for a given learner.

45

This decision may have resulted in a selection bias through differential attrition. However, comparisons were conducted in order to establish the presence of such bias and confirmed that this was not the case.

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 May 2015 (also June 2014) 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.

4.2 Challenges experienced with project evaluation

During the second half of 2014 service delivery protests in the greater Malamulele area in Limpopo resulted in disruption of access to schools for the intervention service delivery organisation (sJsK). After strong efforts afterwards to catch up lost time and effort before the end of the year, it was realised at some point that these would not succeed. It also became clear that the periods of disruption and being barred from going into the area not only made it impossible to complete the 2014 interventions, but would compromise the planned end-line evaluation data. An extension was then granted by the funders to roll over completion of the intervention delivery and final assessments to the first couple of months of 2015. However, protests flared up again and resulted in schools not reopening for six weeks. The further delays implied that the Year 3 interventions intended for the first semester of 2015 had to be abandoned in favour of completing the original Year 2 (2014) interventions, with the funders' permission.

As a result of this, the project team in November 2014 alerted the Research Ethics Committee of the HSRC about these events and the fact that no test administration, quality assurance or data collection would be taking place at the end of 2014. This comprised both a deviation notice, and the regular annual request for renewal of the study. The scheduling change and annual recertification of the study were both approved on 7 December 2014.

4.3 Implications and recommendations

Outcomes are positive after the second 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. In addition, these positive results were obtained even after disruption during the second part of the intervention period and the resultant delay of completion of the intervention. 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.

This report was essentially a replication of the first 2013 report as the intervention was applied in additional schools and included a second home language. The positive results replicated in this report demonstrated that

46

the intervention can be effectively transferred to other similar, low socioeconomic, contexts. Furthermore, the Grade 7 learners were taken from both the first year and the second year schools. The improvements reported for this sample of learners further indicates that the intervention can be effectively up-scaled. Report 1, covering the first cohort over two years, provides a more complete longer-term view.

Similar matters regarding data, their availability and quality, were identified including early and strong coordination between the evaluation and intervention teams to ensure that learner background and achievement data can be linked more easily. This remains an issue due to the inability of schools to provide 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.

Pertaining to data still, an important matter is the one of 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 remained elusive.

Given the experience gained by all of those involved in this stimulating programme, it can be said that the intervention yields a positive impact in its targeted areas for rebuilding language foundations. The intervention was also carefully monitored in its entirety including all the processes, communications, records and data related to ending both the first and now the second year's evaluation cycle. The three reports generated serve to keep all the beneficiaries, participants and other stakeholders informed. 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.

References

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, Achievement, and Inequality: A Seasonal Perspective. *Educational Evaluation and Policy Analysis*, 23(2), 171–191. http://doi.org/10.3102/01623737023002171

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The Effects of Summer Vacation on Achievement Test Scores: A Narrative and Meta-Analytic Review. *Review of Educational Research*, 66(3), 227–268. http://doi.org/10.3102/00346543066003227

Entwisle, D. R., & Alexander, K. L. (1992). Summer setback: Race, poverty, school composition, and mathematics achievement in the first two years of school. *American Sociological Review*, 57(1), 72. http://doi.org/10.2307/2096145

Rambo-Hernandez, K. E., & McCoach, D. B. (2015). High-achieving and average students' reading growth: Contrasting school and summer trajectories. *Journal of Educational Research*, 108(2), 112–129. http://doi.org/10.1080/00220671.2013.850398

47

3. Endnotes

[i] In relation to analysis of the **Grade 1** data:

- For the EGRA Letter Sound test, the value-added model explained 31.5% of the variance in the November scores, with the two standardised (beta) coefficients being 0,233 for the February score and 0,548 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 34,3%, but increased the coefficient for intervention status substantially to 0,741, remaining the largest contribution by far.
- For the EGRA Word Recognition test, the value-added model explained only 3,6% of the variance in the November scores, with the two standardised (beta) coefficients being 0,149 for the February score and 0,125 for intervention status (***p<001). As applied above, the robustness check increased the variance explained by the model only marginally to 8,6%, and increased the coefficient for intervention status slightly to 0,393, becoming the largest contribution by far.
- For the EGRA Non-Word Decoding test, the value-added model explained only 2,7% of the variance in the November scores, with the two standardised (beta) coefficients being 0,147 for the February score (***p<001) and -0,093 for intervention status (*p<05). As applied above, the robustness check increased the variance explained by the model only marginally to 6,6%, and increased the coefficient for intervention status to 0,447, becoming the largest contribution by far.
- For the Oral Pictorial test, the value-added model explained only 2,6% of the variance in the November scores, with the two standardised (beta) coefficients being 0,147 for the February score (***p<001) and 0,089 for intervention status (*p<05). As applied above, the robustness check increased the variance explained by the model to 15,6%, and substantially increased the coefficient for intervention status to 0,633, becoming the largest contribution by far.

[ii] In relation to analysis of the **Grade 4** data:

- For the Schonell Reading Age test, the value-added model explained only 36,0% of the variance in the November scores, with the two standardised (beta) coefficients being 0,569 for the February score and -0,206 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 36,9%, and reduced the coefficient for intervention status to 0,178 (not statistically significant), remaining the second largest contribution behind the February coefficient now decreased a little to 0,559.
- For the EGRA Letter Sound test, the value-added model explained 77,5% of the variance in the November scores, with the two standardised (beta) coefficients being 0,004 for the February score (not statistically significant) and -0,881 for intervention status (***p<001). As applied above, the robustness check increased the variance explained by the model only marginally to 78,1%, and reduced the coefficient for intervention status to 0,765, remaining the largest contribution by far.
- For the EGRA Word Recognition test, the value-added model explained 47,6% of the variance in the November scores, with the two standardised (beta) coefficients being 0,682 for the February score and 0,133 for intervention status (***p<001). As applied above, the robustness check increased the variance explained by the model only marginally to 47,8%, and increased the coefficient for intervention status slightly to 0,142 (not statistically significant), remaining the second largest contribution behind the February coefficient now increased a little to 0,686.
- For the EGRA Non-Word Decoding test, the value-added model explained 34,5% of the variance in the November scores, with the two standardised (beta) coefficients being 0,506 for the February score and 0,312 for intervention status (***p<001). As applied above, the robustness check increased the variance explained by the model only marginally to 34,7%, and increased the coefficient for intervention status slightly to 0,386, remaining the second largest contribution behind the marginally decreased February score coefficient of 0,505.
- For the Literacy 1 (MCQ) sub-test, the value-added model explained 29,9% of the variance in the November scores, with the two standardised (beta) coefficients being 0,090 for the February score (**p<01) and 0,536 for intervention status (***p<001). As applied above, the robustness check increased the variance explained by the model to 34,9%, and decreased the coefficient for intervention status to 0,478, remaining the largest contribution.
- For the Literacy 2 (Sentence Ordering) sub-test, the value-added model explained only 8,6% of the variance in the November scores, with the two standardised (beta) coefficients being 0,097 for the February score (*p<05) and 0,283 for intervention status (***p<001). As applied above, the robustness check increased the variance explained by the model to 11,0%, and also increased the coefficient for intervention status to 0,303 (**p<01), remaining the largest contribution.
- For the Literacy 4 (Open-ended Items) sub-test, the value-added model explained only 26,5% of the variance in the November scores, with the two standardised (beta) coefficients being 0,331 for the February score and 0,350 for intervention status (***p<001). As applied above, the robustness check increased the variance explained by the model to 34,3%, and increased the

48

coefficient for intervention status to 0,474, remaining the largest contribution.

- For the Literacy Total overall test, the value-added model explained 29,5% of the variance in the November scores, with the two standardised (beta) coefficients being 0,288 for the February score and 0,434 for intervention status (***p<001). As applied above, the robustness check increased the variance explained by the model to 36,1%, and increased the coefficient for intervention status slightly to 0,495, remaining the largest contribution.

[iii] In relation to analysis of the **Grade 7** data:

- For the Schonell Reading Age test, the value-added model explained 14,1% of the variance in the November scores, with the two standardised (beta) coefficients being 0,351 for the February score and 0,148 for intervention status (***p<001). Adding monitoring status during February and November and their interaction terms with intervention status in a robustness check did not alter the variance explained by the model, and marginally decreased the coefficient for intervention status to 0,144, remaining the only other large contribution behind the February coefficient now reduced a little to 0,349.
- For the EGRA Letter Sound test, the value-added model explained 68,0% of the variance in the November scores, with the two standardised (beta) coefficients being 0,071 for the February score and 0,815 for intervention status (***p<001). As applied above, the robustness check increased the variance explained by the model only marginally to 68,8%, and reduced the coefficient for intervention status slightly to 0,787, remaining the largest contribution by far.
- For the EGRA Word Recognition test, the value-added model explained 62,7% of the variance in the November scores, with the two standardised (beta) coefficients being 0,789 for the February score and 0,075 for intervention status (***p<001). As applied above, the robustness check did not change the variance explained by the model, but decreased the coefficient for intervention status slightly to 0,073, remaining the second largest contribution behind the February coefficient.
- For the EGRA Non-Word Decoding test, the value-added model explained 35,1% of the variance in the November scores, with the two standardised (beta) coefficients being 0,580 for the February score and 0,123 for intervention status (***p<001). As applied above, the robustness check decreased the variance explained by the model only marginally to 35,0%, and decreased the coefficient for intervention status slightly to 0,120, remaining the second largest contribution behind the February score coefficient.
- For the Literacy 1 (MCQ) sub-test, the value-added model explained only 12,8% of the variance in the November scores, with the two standardised (beta) coefficients being 0,303 for the February score and -0,199 for intervention status (***p<001). As applied above, the robustness check did not change the variance explained by the model, and decreased the coefficient for intervention status marginally to 0,198, remaining the second largest contribution behind the February score coefficient.
- For the Literacy 3 (Paragraph Writing) sub-test, the value-added model explained only 15,7% of the variance in the November scores, with the two standardised (beta) coefficients being 0,343 for the February score and 0,197 for intervention status (***p<001). As applied above, the robustness check increased the variance explained by the model slightly to 15,8%, and increased the coefficient for intervention status to 0,207, remaining the second largest contribution behind the February score coefficient.
- For the Literacy 4 (Open-ended Items) sub-test, the value-added model explained 16,8% of the variance in the November scores, with the two standardised (beta) coefficients being 0,367 for the February score and -0,166 for intervention status (***p<001). As applied above, the robustness check increased the variance explained by the model to 17,6%, and decreased the coefficient for intervention status to 0,145, remaining the second largest contribution behind the February score coefficient.
- For the Literacy Total overall test, the value-added model explained 40,0% of the variance in the November scores, with the two standardised (beta) coefficients being 0,582 for the February score and 0,230 for intervention status (***p<001). As applied above, the robustness check increased the variance explained by the model to 40,1%, and decreased the coefficient for intervention status slightly to 0,225, remaining the second largest contribution behind the February score coefficient.

49

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	***17.3	10.8	*64.5	60.3	***22.9	13.0	30.6	28.8
	(s.e.)	(0.7)	(0.9)	(1.0)	(1.5)	(1.4)	(1.5)	(1.4)	(2.4)
	SD	17.1	14.9	24.4	24.8	16.8	13.2	17.4	19.5
Early Grade Reading Assessment (EGRA) Word Recognition	Mean	0.2	0.1	**39.0	31.9	0.2	0.0	**29.7	19.1
	(s.e.)	(0.1)	(0.0)	(1.3)	(1.8)	(0.2)	(0.0)	(2.1)	(3.3)
	SD	1.8	0.5	31.0	28.5	2.0	0.0	26.0	26.7
Early Grade Reading Assessment (EGRA) Non-word Decoding	Mean	0.5	0.5	50.1	43.4	**0.8	0.0	***43.0	27.7
	(s.e.)	(0.1)	(0.3)	(1.5)	(2.0)	(0.5)	(0.0)	(2.4)	(3.9)
	SD	3.3	4.4	33.8	33.1	6.7	0.0	28.9	31.7
Oral Pictorial	Mean	***18.4	14.3	*75.8	70.8	18.0	14.8	64.1	60.7
	(s.e.)	(0.8)	(0.9)	(0.9)	(1.5)	(1.2)	(1.4)	(1.7)	(2.7)
	SD	17.5	16.3	20.5	23.9	14.9	12.5	21.2	22.0

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

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

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	***6.2	5.7	***7.5	7.0	6.2	5.9	***6.9	6.4
	(s.e.)	(0.1)	(0.1)	(0.1)	(0.1)	(0.1)	(0.2)	(0.1)	(0.2)
	SD	1.4	1.7	1.2	1.5	1.1	1.5	0.7	1.3
Early Grade Reading Assessment (EGRA) Letter Sound	Mean	16.5	19.4	***84.0	74.6	17.6	19.0	10.3	13.0
	(s.e.)	(0.8)	(1.3)	(0.8)	(1.8)	(1.3)	(2.4)	(0.7)	(2.5)
	SD	17.7	17.7	18.4	25.2	15.8	15.3	8.7	16.1
Early Grade Reading Assessment (EGRA) Word Recognition	Mean	***58.0	47.4	***84.7	77.9	60.1	51.3	***79.0	64.3
	(s.e.)	(1.3)	(2.2)	(0.9)	(1.9)	(2.2)	(4.7)	(1.8)	(4.4)
	SD	29.2	30.0	20.8	25.7	26.4	29.5	21.9	28.5
Early Grade Reading Assessment (EGRA) Non-word Decoding	Mean	***61.0	51.0	***85.2	78.2	*62.1	53.8	**71.5	62.3
	(s.e.)	(1.1)	(2.0)	(0.8)	(1.7)	(1.8)	(4.2)	(1.4)	(3.8)
	SD	24.7	27.4	18.4	23.5	22.2	26.3	17.3	24.4
Literacy 1: Multiple choice Items	Mean	22.0	24.5	*77.0	70.5	18.2	18.0	*31.3	21.4
	(s.e.)	(1.2)	(2.1)	(1.4)	(2.5)	(2.0)	(4.2)	(2.4)	(3.9)
	SD	27.1	28.2	30.3	32.6	24.9	25.6	29.0	25.3
Literacy 2: Sentence Ordering	Mean	14.9	11.7	48.9	44.8	15.4	13.5	25.3	22.0
	(s.e.)	(0.9)	(1.3)	(1.6)	(2.5)	(1.5)	(3.2)	(2.1)	(4.1)
	SD	20.4	16.9	36.4	35.2	18.8	19.2	25.4	26.6
Literacy 4: Open-ended Items	Mean	*9.0	5.7	*42.6	35.9	4.0	2.9	**11.6	3.7
	(s.e.)	(0.7)	(0.9)	(1.4)	(2.3)	(0.7)	(1.1)	(1.5)	(1.2)
	SD	15.0	11.6	32.4	30.5	8.5	6.7	18.6	7.9
Literacy: Total Score	Mean	14.3	12.3	*51.6	46.0	9.9	8.7	*19.2	12.1
	(s.e.)	(0.9)	(2.1)	(1.2)	(2.0)	(0.8)	(1.5)	(1.5)	(1.6)
	SD	19.4	26.5	27.8	26.3	9.3	9.1	17.8	10.7

* 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	7.3	7.2	8.0	7.9	7.5	7.2	*7.6	7.2
	(s.e.)	(0.0)	(0.1)	(0.0)	(0.1)	(0.2)	(0.2)	(0.0)	(0.3)
	SD	1.1	1.2	1.3	1.4	3.5	1.2	1.1	1.6
Early Grade Reading Assessment (EGRA) Letter Sound	Mean	13.8	*18.2	85.8	87.0	9.8	8.3	25.3	21.0
	(s.e.)	(0.6)	(2.1)	(0.5)	(1.2)	(0.7)	(1.6)	(1.1)	(3.5)
	SD	21.5	27.2	17.2	13.8	14.9	12.3	23.3	20.0
Early Grade Reading Assessment (EGRA) Word Recognition	Mean	***81.4	74.6	**90.0	85.4	***82.0	71.5	87.5	84.0
	(s.e.)	(0.7)	(2.1)	(0.5)	(1.8)	(0.9)	(3.9)	(0.8)	(3.2)
	SD	22.1	26.5	17.2	20.4	19.8	30.3	17.0	17.9
Early Grade Reading Assessment (EGRA) Non-word Decoding	Mean	*72.8	69.0	82.9	80.1	*73.0	66.1	78.1	73.0
	(s.e.)	(0.6)	(1.9)	(0.6)	(1.8)	(0.9)	(3.3)	(0.8)	(3.7)
	SD	21.4	24.8	18.4	20.5	19.6	25.4	17.7	20.7
Literacy 1: Multiple choice Items	Mean	40.0	41.7	66.0	64.8	41.1	38.4	53.8	45.5
	(s.e.)	(0.9)	(2.7)	(0.8)	(2.4)	(1.4)	(4.1)	(1.4)	(6.4)
	SD	31.0	33.1	26.9	27.2	30.1	29.1	31.2	30.1
Literacy 3: Writing exercise	Mean	21.3	19.8	33.2	32.6	**20.7	10.8	**20.6	9.9
	(s.e.)	(0.8)	(2.1)	(0.9)	(2.6)	(1.1)	(2.7)	(1.1)	(2.2)
	SD	25.6	25.7	29.4	32.5	24.2	19.2	23.8	15.5
Literacy 4: Open-ended Items	Mean	13.7	12.6	18.2	18.8	11.6	9.1	11.2	8.3
	(s.e.)	(0.5)	(1.3)	(0.5)	(1.5)	(0.7)	(1.9)	(0.7)	(2.4)
	SD	16.7	16.0	17.7	18.0	16.3	13.9	14.8	11.2
Literacy: Total Score	Mean	23.3	22.4	35.7	34.1	22.1	18.3	25.8	19.4
	(s.e.)	(0.5)	(1.3)	(0.5)	(1.6)	(0.7)	(2.1)	(0.8)	(2.6)
	SD	15.9	15.9	17.6	17.7	16.2	15.0	17.1	12.4

* 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**	**Feb 2014**	**Feb → May**	**May 2015**	**Feb 2014**	**Feb → May**	**May 2015**
EGRA Letter Sound	**Feb 2014**	10.8 (93) 11.1	50.6 (93) 26.3	61.4 (93) 26.8	***18.6 (446) 17.8	46.5 (446) 26.1	65.1 (446) 23.8
	May 2015	19.2 (104) 14.9	***54.8 (104) 26.6	***74.0 (104) 23.6	16.8 (435) 17.6	45.4 (435) 25.8	62.2 (435) 24.0
	Either	12.3 (197) 13.9	***52.8 (197) 26.5	**68.1 (197) 25.9	*18.4 (342) 18.6	44.0 (342) 25.5	62.4 (342) 23.2
EGRA Word Recognition	**Feb 2014**	0.2 (93) 1.7	39.2 (93) 32.7	39.5 (93) 33.0	0.3 (447) 1.9	38.7 (447) 30.5	38.9 (447) 30.6
	May 2015	0.1 (104) 1.0	***50.4 (104) 33.5	***50.5 (104) 33.5	0.3 (436) 2.0	36.0 (436) 29.5	36.3 (436) 29.8
	Either	0.2 (197) 1.4	***45.1 (197) 33.5	***45.3 (197) 33.6	0.3 (343) 2.0	35.1 (343) 28.6	35.4 (343) 28.8
EGRA Non-Word Decoding	**Feb 2014**	0.6 (92) 3.7	49.2 (92) 36.5	49.8 (92) 36.9	0.5 (443) 3.2	49.7 (443) 32.9	50.2 (443) 33.2
	May 2015	0.2 (102)	**57.3 (102) 33.7	*57.5 (102) 33.7	0.6 (433) 3.5	47.8 (433) 33.2	48.4 (433) 33.7
	Either	0.4 (194) 2.9	*53.5 (194) 35.2	53.8 (194) 35.4	0.6 (341) 3.5	47.4 (341) 32.4	48.0 (341) 32.8
Oral Pictorial	**Feb 2014**	16.6 (93) 19.3	49.9 (93) 28.8	66.5 (93) 25.0	18.8 (446) 17.1	***58.9 (446) 22.1	***77.7 (446) 18.9
	May 2015	19.5 (103) 18.1	60.8 (103) 23.4	*80.3 (103) 20.0	18.2 (436) 17.4	56.5 (436) 23.6	74.7 (436) 20.5
	Either	18.1 (196) 18.7	55.6 (196) 26.6	73.7 (196) 23.5	18.6 (343) 16.9	58.3 (343) 21.7	76.9 (343) 18.5
CONTROL SCHOOLS							
Sub-test	**Monitor date**	**Feb 2014**	**Feb → May**	**May 2015**	**Feb 2014**	**Feb → May**	**May 2015**
EGRA Letter Sound	**Feb 2014**	26.1 (19) 17.9	7.5 (19) 25.1	33.6 (19) 20.4	22.4 (130) 16.7	7.8 (130) 18.5	30.2 (130) 16.9
	May 2015	*24.7 (110) 16.5	9.0 (110) 19.1	***33.6 (110) 15.8	17.9 (39) 16.9	4.3 (39) 19.9	22.2 (39) 19.0
	Either	***24.9 (129) 16.6	8.8 (129) 20.0	***33.6 (129) 16.5	10.0 (20) 11.7	1.3 (20) 13.3	11.3 (20) 8.4
EGRA Word Recognition	**Feb 2014**	0.0 (19) 0.0	21.6 (19) 23.9	21.6 (19) 23.9	0.2 (130) 2.1	30.7 (130) 25.8	30.9 (130) 26.2
	May 2015	0.2 (110) 2.3	***35.2 (110) 25.1	***35.4 (110) 25.5	0.0 (39) 0.0	13.5 (39) 19.9	13.5 (39) 19.9
	Either	0.2 (129) 2.1	***33.2 (129) 25.3	***33.4 (129) 25.7	0.0 (20) 0.0	5.8 (20) 11.0	5.8 (20) 11.0
EGRA Non-Word Decoding	**Feb 2014**	0.0 (19) 0.0	33.1 (19) 31.1	33.1 (19) 31.1	0.9 (130) 7.2	43.6 (130) 27.9	44.5 (130) 28.4
	May 2015	1.0 (110) 7.8	***49.4 (110) 24.5	***50.4 (110) 24.8	0.0 (39) 0.0	22.1 (39) 29.5	22.1 (39) 29.5
	Either	0.9 (129) 7.2	***47.0 (129) 26.1	***27.9 (129) 26.4	0.0 (20) 0.0	11.7 (20) 24.4	11.7 (20) 24.4
Oral Pictorial	**Feb 2014**	21.1 (19) 15.2	40.0 (19) 20.8	61.1 (19) 18.2	17.6 (129) 14.9	46.9 (129) 23.6	64.5 (129) 21.6
	May 2015	18.2 (110) 14.8	***50.0 (110) 27.7	***68.2 (110) 19.7	17.6 (38) 15.5	34.5 (38) 21.3	52.1 (38) 21.1
	Either	18.6 (129) 14.8	***48.5 (129) 22.6	***67.1 (129) 19.6	14.2 (19) 15.4	28.9 (19) 20.8	43.2 (19) 20.3

* 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

Figures in each cell = Mean % (N) SD		Monitored			Not monitored		
(1st four subtests) PROJECT SCHOOLS							
Sub-test	**Monitor date**	**Feb 2014**	**Feb → May**	**May 2015**	**Feb 2014**	**Feb → May**	**May 2015**
Schonell Reading Age (age)	**Feb 2014**	6.1 (79) 1.4	1.4 (79) 1.1	7.5 (79) 1.0	6.2 (423) 1.4	1.3 (423) 1.2	7.5 (423) 1.3
	May 2015	**6.6 (83) 1.5	1.4 (83) 1.2	***8.0 (83) 1.2	6.1 (419) 1.4	1.3 (419) 1.2	7.4 (419) 1.2
	Either	6.3 (162) 1.5	1.4 (162) 1.1	***7.7 (162) 1.1	6.1 (340) 1.3	1.3 (340) 1.2	7.4 (340) 1.2
EGRA Letter Sound	**Feb 2014**	10.1 (79) 9.2	***80.3 (79) 14.8	***90.4 (79) 11.3	***17.7 (425) 18.7	65.1 (425) 26.6	83.0 (425) 19.2
	May 2015	***25.5 (83) 24.9	61.0 (83) 27.9	86.5 (83) 16.7	14.7 (421) 15.4	*68.7 (421) 25.1	83.5 (421) 18.7
	Either	18.0 (162) 20.4	70.4 (162) 24.4	***88.4 (162) 14.4	15.8 (342) 16.3	66.1 (342) 26.2	81.9 (342) 19.7
EGRA Word Recognition	**Feb 2014**	59.0 (79) 28.9	29.5 (79) 22.9	88.5 (79) 17.3	57.8 (428) 29.3	26.2 (428) 21.2	84.0 (428) 21.3
	May 2015	***68.4 (83) 29.8	19.2 (83) 17.8	87.6 (83) 21.2	56.0 (424) 28.7	***28.2 (424) 21.8	84.1 (424) 20.7
	Either	**63.8 (162) 29.7	24.2 (162) 21.1	*88.0 (162) 19.3	55.3 (345) 28.7	27.8 (345) 21.6	83.1 (345) 21.3
EGRA Non-Word Decoding	**Feb 2014**	62.2 (79) 26.6	26.2 (79) 22.9	88.4 (79) 17.1	60.7 (425) 24.3	23.8 (425) 21.3	84.6 (425) 18.6
	May 2015	**67.4 (83) 26.0	19.6 (83) 19.5	87.0 (83) 18.1	59.7 (421) 24.2	*25.1 (421) 21.8	84.8 (421) 18.5
	Either	*64.9 (162) 26.3	22.8 (162) 21.4	*87.7 (162) 17.5	59.1 (342) 23.7	24.9 (342) 21.6	84.0 (342) 18.7
(1st four subtests) CONTROL SCHOOLS							
Sub-test	**Monitor date**	**Feb 2014**	**Feb → May**	**May 2015**	**Feb 2014**	**Feb → May**	**May 2015**
Schonell Reading Age (age)	**Feb 2014**	6.1 (69) 1.5	0.8 (69) 1.2	7.0 (69) 0.7	6.3 (81) 0.7	0.6 (81) 0.5	6.9 (81) 0.6
	May 2015	**6.4 (109) 0.7	0.6 (109) 0.5	7.0 (109) 0.7	5.8 (41) 1.7	**1.0 (41) 1.5	6.8 (41) 0.7
	Either	6.1 (122) 1.2	0.8 (122) 1.0	6.9 (122) 0.6			
	Both	*6.7 (28) 0.6	0.5 (28) 0.5	*7.2 (28) 0.7	-	-	-
EGRA Letter Sound	**Feb 2014**	11.9 (69) 10.9	**-2.8 (69) 11.6	9.1 (69) 6.7	***22.4 (81) 17.7	-11.1 (81) 21.2	11.3 (81) 10.0
	May 2015	17.9 (109) 17.1	-8.2 (109) 19.1	9.8 (109) 9.4	16.7 (41) 11.7	-5.1 (41) 14.1	11.6 (41) 6.5
	Either	***20.5 (122) 16.1	-9.1 (122) 19.2	***11.4 (122) 9.0			
	Both	4.9 (28) 3.3	*0.4 (28) 5.1	5.4 (28) 5.1	-	-	-
EGRA Word Recognition	**Feb 2014**	62.2 (69) 26.9	17.4 (69) 18.0	79.7 (69) 22.2	58.3 (81) 26.0	20.2 (81) 19.2	78.5 (81) 21.7
	May 2015	62.2 (109) 25.6	18.2 (109) 18.1	80.4 (109) 20.5	54.5 (41) 28.0	20.8 (41) 20.0	75.3 (41) 25.0
	Either	57.0 (122) 26.6	*20.4 (122) 19.4	77.4 (122) 22.8			
	Both	**73.6 (28) 21.0	12.4 (28) 13.4	86.0 (28) 15.9	-	-	-
EGRA Non-Word Decoding	**Feb 2014**	61.0 (69) 26.4	12.2 (69) 20.9	73.2 (69) 19.2	63.0 (81) 18.1	7.0 (81) 18.7	70.1 (81) 15.6
	May 2015	*64.8 (109) 18.7	7.3 (109) 18.3	72.1 (109) 16.7	54.9 (41) 28.7	*15.0 (41) 22.6	69.9 (41) 19.1
	Either	60.3 (122) 22.4	9.7 (122) 20.3	70.0 (122) 16.7			
	Both	*70.0 (28) 19.8	8.0 (28) 17.7	*77.9 (28) 18.6	-	-	-

* 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 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	Feb 2014	Feb → May	May 2015	Feb 2014	Feb → May	May 2015
Literacy 1: MCQ	Feb 2014	21.3 (79) 25.6	**65.2 (79) 32.5	**86.5 (79) 21.7	22.1 (418) 27.4	53.1 (418) 39.3	75.2 (418) 31.3
	May 2015	26.2 (84) 31.1	***67.9 (84) 30.8	***94.0 (84) 14.8	21.1 (413) 26.2	52.4 (413) 39.4	73.5 (413) 31.4
	Either	23.8 (163) 28.6	***66.6 (163) 31.6	***90.4 (163) 18.8	21.1 (334) 26.3	49.4 (334) 40.3	70.4 (334) 32.6
Literacy 2: Sentence Ordering	Feb 2014	13.6 (81) 20.8	36.4 (81) 43.6	50.0 (81) 37.9	15.2 (408) 20.3	33.5 (408) 39.7	48.7 (408) 36.1
	May 2015	***24.1 (84) 20.7	40.8 (84) 39.0	***64.9 (84) 35.2	13.0 (405) 19.8	32.6 (405) 40.5	45.6 (405) 35.8
	Either	**18.9 (165) 21.4	38.6 (165) 41.2	***57.6 (165) 37.2	12.9 (324) 19.6	31.6 (324) 39.7	44.5 (324) 35.2
Literacy 4: Open-ended Items	Feb 2014	6.9 (81) 9.4	29.3 (81) 29.1	36.2 (81) 29.2	9.4 (434) 15.8	34.5 (434) 30.3	43.8 (434) 32.8
	May 2015	***14.9 (84) 24.2	***55.1 (84) 30.7	***70.0 (84) 31.6	7.8 (431) 12.2	29.5 (431) 28.2	37.3 (431) 29.7
	Either	*10.9 (165) 18.9	***42.5 (165) 32.5	***53.4 (165) 34.8	8.0 (350) 12.8	29.5 (350) 28.1	37.6 (350) 29.9
Literacy: Total	Feb 2014	11.8 (81) 9.2	38.1 (81) 25.2	49.9 (81) 24.8	14.7 (436) 20.7	37.2 (436) 28.8	51.9 (436) 28.3
	May 2015	**19.6 (84) 20.1	***53.7 (84) 25.6	***73.3 (84) 25.5	13.2 (433) 19.1	34.2 (433) 27.7	47.4 (433) 26.2
	Either	15.8 (165) 16.2	***46.0 (165) 26.5	***61.8 (165) 27.6	13.6 (352) 20.7	33.2 (352) 28.2	46.8 (352) 26.5
		(final four subtests) CONTROL SCHOOLS					
Sub-test	Monitor date	Feb 2014	Feb → May	May 2015	Feb 2014	Feb → May	May 2015
Literacy 1: MCQ	Feb 2014	14.5 (69) 23.2	15.5 (69) 37.3	30.0 (69) 29.8	21.4 (81) 26.0	11.1 (81) 34.6	32.5 (81) 28.4
	May 2015	18.7 (109) 24.2	13.5 (109) 34.6	32.1 (109) 29.7	17.1 (41) 27.0	12.2 (41) 39.3	29.3 (41) 27.1
	Either	19.9 (122) 26.3	11.5 (122) 36.1	31.4 (122) 27.9	-	-	-
	Both	10.7 (28) 15.9	20.2 (28) 34.4	31.0 (28) 33.9			
Literacy 2: Sentence Ordering	Feb 2014	14.3 (69) 19.2	11.4 (69) 27.7	25.7 (69) 24.6	16.4 (81) 18.6	8.6 (81) 28.0	25.0 (81) 26.2
	May 2015	17.0 (109) 19.5	9.2 (109) 27.8	26.1 (109) 26.4	11.3 (41) 16.5	11.9 (41) 28.1	23.2 (41) 22.6
	Either	14.7 (122) 18.0	9.7 (122) 28.0	24.4 (122) 25.0	-	-	-
	Both	18.8 (28) 22.2	10.7 (28) 27.6	29.5 (28) 27.3			
Literacy 4: Open-ended	Feb 2014	4.6 (69) 8.0	6.2 (69) 19.9	10.8 (69) 16.7	3.5 (81) 8.9	8.8 (81) 18.7	12.3 (81) 20.2
	May 2015	3.0 (109) 8.0	***10.6 (109) 18.6	*13.6 (109) 19.5	*6.7 (41) 9.3	-0.5 (41) 18.7	6.3 (41) 14.9
	Either	4.6 (122) 9.2	5.7 (122) 19.1	10.2 (122) 18.8	-	-	-
	Both	1.6 (28) 4.0	*15.8 (28) 17.7	17.4 (28) 17.1			
Literacy: Total	Feb 2014	9.1 (69) 8.2	9.5 (69) 17.5	18.6 (69) 16.5	10.5 (81) 10.2	9.2 (81) 17.0	19.7 (81) 18.9
	May 2015	9.9 (109) 9.6	10.8 (109) 17.1	20.6 (109) 18.7	9.9 (41) 8.5	5.4 (41) 17.0	15.4 (41) 14.5
	Either	10.3 (122) 9.6	8.0 (122) 17.0	18.3 (122) 17.7	-	-	-
	Both	8.0 (28) 7.7	*15.4 (28) 16.8	23.3 (28) 18.1			

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

Figures in each cell = Mean % (N) SD		Monitored			Not monitored		
		(1st four subtests) PROJECT SCHOOLS					
Sub-test	**Monitor date**	**Feb 2014**	**Feb → May**	**Nov 2014**	**Feb 2014**	**Feb → May**	**Nov 2014**
Schonell Reading Age (age)	**Feb 2014**	7.2 (109) 1.0	0.7 (109) 0.8	7.8 (109) 1.3	7.3 (993) 1.2	0.7 (993) 1.0	8.0 (993) 1.3
EGRA Letter Sound	**Feb 2014**	7.0 (109) 7.6	***79.7 (109) 19.4	86.7 (109) 16.9	***14.6 (991) 22.4	71.1 (991) 27.1	85.7 (991) 17.2
EGRA Word Recognition	**Feb 2014**	81.8 (109) 18.9	8.3 (109) 10.6	90.1 (109) 15.1	81.3 (993) 22.4	8.6 (993) 13.4	90.0 (993) 17.5
EGRA Non-Word Decoding	**Feb 2014**	72.4 (109) 18.2	10.5 (109) 15.7	82.8 (109) 16.6	72.8 (994) 21.7	10.1 (994) 17.9	82.9 (994) 18.6
		(1st four subtests) CONTROL SCHOOLS					
Sub-test	**Monitor date**	**Feb 2014**	**Feb → May**	**Nov 2014**	**Feb 2014**	**Feb → May**	**Nov 2014**
Schonell Reading Age (age)	**Feb 2014**	7.0 (50) 0.7	0.3 (50) 0.7	7.4 (50) 1.0	7.5 (420) 3.7	0.1 (420) 3.6	*7.7 (420) 1.1
EGRA Letter Sound	**Feb 2014**	6.1 (50) 4.2	2.5 (50) 10.5	8.5 (50) 9.7	10.3 (421) 15.7	***17.0 (421) 24.8	***27.3 (421) 23.6
EGRA Word Recognition	**Feb 2014**	78.5 (50) 19.6	6.1 (50) 13.7	84.6 (50) 17.1	82.4 (421) 19.8	5.5 (421) 12.9	87.8 (421) 17.0
EGRA Non-Word Decoding	**Feb 2014**	71.6 (50) 22.1	4.6 (50) 20.9	76.2 (50) 18.8	73.2 (421) 19.3	5.1 (421) 18.2	78.3 (421) 17.5

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

Annexure 6b: Comparative monitoring statistics for Grade 7 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**	**Feb 2014**	**Feb → May**	**Nov 2014**	**Feb 2014**	**Feb → May**	**Nov 2014**
Literacy 1: MCQ	**Feb 2014**	28.9 (109) 27.2	31.3 (109) 31.5	60.2 (109) 29.9	***41.2 (993) 31.1	25.4 (993) 34.9	*66.6 (993) 26.5
Literacy 3: Paragraph Writing	**Feb 2014**	16.3 (106) 28.9	11.2 (106) 32.3	27.5 (106) 25.4	*21.9 (972) 25.2	11.9 (972) 32.0	*33.8 (972) 29.8
Literacy 4: Open-ended Items	**Feb 2014**	*17.5 (109) 19.5	0.4 (109) 19.4	17.8 (109) 19.1	13.2 (983) 16.3	*4.9 (983) 19.7	18.2 (983) 17.5
Literacy: Total	**Feb 2014**	21.5 (109) 18.7	11.3 (109) 16.1	32.8 (109) 19.1	23.6 (983) 15.4	12.5 (993) 15.5	35.9 (983) 17.5
		(final four subtests) CONTROL SCHOOLS					
Sub-test	**Monitor date**	**Feb 2014**	**Feb → May**	**Nov 2014**	**Feb 2014**	**Feb → May**	**Nov 2014**
Literacy 1: MCQ	**Feb 2014**	21.5 (50) 22.6	*23.8 (50) 29.5	45.3 (50) 31.4	***43.3 (431) 30.1	11.4 (431) 35.4	*54.8 (431) 31.1
Literacy 3: Paragraph Writing	**Feb 2014**	10.9 (23) 21.1	9.2 (23) 29.7	20.1 (23) 23.5	*21.2 (427) 24.2	-0.6 (427) 25.5	20.6 (427) 23.9
Literacy 4: Open-ended	**Feb 2014**	***19.4 (50) 19.4	-13.4 (50) 17.6	6.0 (50) 9.6	10.7 (431) 16.7	***1.1 (431) 15.3	**11.8 (431) 15.2
Literacy: Total	**Feb 2014**	17.2 (50) 13.9	2.1 (50) 11.7	19.2 (50) 13.9	*22.7 (431) 16.3	3.8 (431) 14.6	**26.5 (431) 17.3

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