JumpStart Report on Foundation Phase Mathematics outcomes 2016-2019

Michael & Susan Dell FOUNDATION

JumpStart

This evaluation is commissioned by JumpStart and funded by the Michael and Susan Dell Foundation.

Date of publication:
30 October 2020

Suggested citation:
Moloi, Q., & Roberts, N. (2020) *JumpStart learning outcomes evaluation 2016-2020,* Kelello: Johannesburg

Enquiries to:
Nicky Roberts
Director: Kelello
E-mail: nicky@kelello.org;
Twitter: @Kelello_
Mobile: +27 71 525 8389

Table of Contents

List of tables

List of figures

Introduction

Children who lack educational support at home and also do not get a good grasp of what they are taught in school are most likely to leave school early. Laying solid foundational skills of literacy and numeracy do not only give children confidence to continue participating in schooling, but also provides them with required skills to engage in further learning. Various studies have shown that South African children's performance, in mathematics in particular, is much lower than in other African countries of a lower socio-economic status (SES) and that only about seven percent of South African grade 6 learners from lower SES groups achieve "Advanced" levels of mathematics (SACMEQ, 2017).

Some of the initiatives taken by the government to address these challenges include providing bursaries for aspiring graduates to study at university following teaching of primary school mathematics as a career of choice. The Department of Education also provides materials and workbooks that are intended to consolidate the teaching of languages and mathematics, particularly at the Foundation Phase. Increasingly the government is encouraging evidence-based interventions for either improvement or policy formulation. However, most of the interventions in schooling have focused at the Further Education and Training (FET) level, particularly at grades 10-12. Despite all the initiatives, participation in mathematics and success rates in the subject continue to be a matter of concern.

This report covers the intervention made by the Jumpstart Programme in the Ekurhuleni Education District in Gauteng between 2016 and 2019. The focus of this report is on whether the JumpStart intervention has made any impact in terms of improving learner skills and competencies in early mathematics at the Foundation Phase (grades 1-3). The report presents the outcomes of the analysis done on data that was collected in schools using the Early Grade Mathematics Assessment (EGMA) tool specially designed for both summative and formative assessments. The analysis of learning outcomes covers the period from the inception of the intervention in 2016 to 2019.

The purpose of this analysis was to explore the data, establish emerging patterns in the impact of the intervention and begin to suggest, based on emerging patterns, what the next steps could be in terms of sustaining and possibly expanding the intervention to scale. This stage of the analysis is broad and at a very high level mainly for decision-makers.

This report will be complemented by a qualitative report which will provide more details on the perceptions of value and lessons learnt and recommendations for further improvement of the intervention design, from those directly involved in the JumpStart intervention: interns (teaching assistants); Foundation Phase teachers, Heads of department and school principals.

Project setting

Over four years since 2016, the JumpStart Programme has managed and sustained an intervention to improve the teaching and learning of mathematics with a focus on the Foundation Phase (FP) in the District of Ekurhuleni South in Gauteng. The Ekhuruleni South District services schools in a predominantly township area marked by dense populations of mainly working-class people, many of whom with either no employment at all or with some unstable employment. Unemployment and associated socio-economic problems in this District and others which have mushroomed in the previously mining areas in Gauteng, often result in children and youth dropping out of school to either join the unemployed working class or swell the numbers of out-of-school children and youth in these areas. The district is described by the Department of Basic Education as being "urban with some informal settlements" (Department of Basic Education, 2015). Predominant home languages in this area are: English (30%), IsiZulu (26%), Sesotho (14%) and Afrikaans (14%). Relative to other districts in South Africa, Ekurhuleni South is small in geographic area, caters for a large number of learners and has a very high population density. This is a relatively well performing district, within one of the top performing provinces in South Africa. Most the schools in this district are classified as quintile 3[1], with 52% of the schools being no-fee schools (Department of Basic Education, 2015).

[1] South African public schools are classified into quintiles with quintile 1 reflecting schools in the most under-resources communities, and quintile 5 schools being in the most affluent communities. Pro-poor policies legislate that parents do not pay schools for children in quintiles 1-3 (no fee schools), and these schools receive the greatest subsidy per child.

Project design

The primary inputs in the JumpStart intervention are the Number Sense workbooks, use of which is supported by school interns visiting each class twice weekly (Roberts, 2020). In Foundation Phase (Grades R-3) in South Africa mathematics is taught in a wide range of languages. The Number Sense workbooks are available in the following languages: Afrikaans, English, Sepedi, Sesotho, Xitsonga, Setswana, isiXhosa and isiZulu. JumpStart began using paper-based workbooks in five Ekurhuleni South schools in 2016. To facilitate the smooth running of these classroom sessions, JumpStart's project delivery manager and two team leaders engage in both formalised workshops and informal mentorship to support teachers and interns in the classroom. The interns in the schools are further supported with school-based visits by a JumpStart project director.

As explained in Roberts (2020) learners experience the JumpStart intervention as having two mathematics lessons per week where the normal class teacher is supported by an intern and they work through a Number Sense workbook. The workbook is graded to the learners' level of development. The Jumpstart intern moves around the class with a tablet application and mark the learners' work.

The inclusion of interns in the programme has an important secondary benefit in the South African context of high youth unemployment (Roberts, 2020). This is seen as the programme's biggest community input. These interns are previously unemployed young South African adults who are employed for a year or two to perform mathematics tutoring in the classroom. Through the internship they gain valuable employment experience, develop a new skill set, and in some cases, find a path to higher education. In addition, the monthly allowance for these interns is subsidized by public youth employment funds. In this way two problems are addressed simultaneously: youth unemployment and lack of work experience, and additional capacity in schools to assist teachers.

When reflecting on the approach to mathematics, the JumpStart intervention is detailed and coherent as it builds on the research base and instructional design of the Number Sense Mathematics Programme. This has been created to support children's development of a "robust sense of number and a deep understanding of mathematics" (Brombacher and Associates, 2019). The materials consist of 12 workbooks (four for each grade level in Foundation Phase). Each workbook has 48 pages and focuses on the development of key mathematical concepts. The workbooks are not explicitly linked to the curriculum but rather designed to complement the curriculum focusing on key skill development in:

- working fluently and flexibly with numbers and number concepts;
- a rich understanding of the meaning of number;
- a wide range of effective strategies for solving a large variety of number problems.

The development of these skills can be broken into three types of activities: counting, manipulating numbers, and solving problems. In the grade 1 workbooks, counting has a majority focus, but by grade 3, the focus has shifted predominantly to manipulating numbers and solving problems.

Project roll out

Since its inception in 2016 the JumpStart Project has grown in three Phases. The changes in scope and scale of the JumpStart intervention were as follows:

- Phase 1 commenced in 2016 with 5 schools;
- Phase 2 commenced in 2017 with another 5 schools; and
- Phase 3 commenced in 2018 with another 10 schools.

In total there were 20 schools sampled for intervention (Treatment) from Ekurhuleni Education District and 12 schools, from the same district, were included for control purposes.

The phases in which schools got into the Jumpstart Project and also the data collection points in the period 2016-2019 are summarised in Table 1. Baseline data on the cohort was collected during the first term in the year that the schools started participating in the Project (Y_0) and then at the same time in each of the years of participation, i.e. Y_1, Y_2 and Y_3.

Table 1: JumpStart intervention data collection points: 2016-2019

2016 Phase 1 (5 schools)	**2017 Phase 2 (5+5 schools)**	**2018 Phase 3 (5+5+10 schools + control schools)**	**2019**
Grade 1 (Y_0)	Grade 1 (Y_1)	Grade 1 (Y_2)	Grade 1 (Y_3)
Grade 2 (Y_0)	Grade 2 (Y_1)	Grade 2 (Y_2)	Grade 2 (Y_3)
Grade 3 (Y_0)	Grade 3 (Y_1)	Grade 3 (Y_2)	Grade 3 (Y_3)
	Grade 1 (Y_0)	Grade 1 (Y_1)	Grade 1 (Y_2)
	Grade 2 (Y_0)	Grade 2 (Y_1)	Grade 2 (Y_2)
	Grade 3 (Y_0)	Grade 3 (Y_1)	Grade 3 (Y_2)
			Grade 4 (Y_3)
		Cntrl Grade 1 (Y_0)	Cntrl Grade 1 (Y_1)
		Cntrl Grade 2 (Y_0)	Cntrl Grade 2 (Y_2)
		Cntrl Grade 3 (Y_0)	Cntrl Grade 3 (Y_3)
			Cntrl Grade 1 (Y_0)
			Cntrl Grade 2 (Y_0)
			Cntrl Grade 3 (Y_0)

Altogether 32 primary schools participated either as Treatment or Control schools in the JumpStart intervention between 2016 and 2019. For the purposes of this study, we focus our analysis on 19 of the 20 treatment schools (as the data for one school was incomplete). Of the 12 control schools involved in the intervention, we focus on the 8 control schools that were classified as quintile 1-4 schools, the remaining 4 control schools are excluded from the sample as they are quintile 5.

Research design

The Jumpstart Programme took a quasi-experimental cross-sectional study design. Treatment schools received support and data on performance was obtained from random selections of learners across grades 1 to 3 during the first term of the school year. In the Control schools (all in Ekhuruleni South district) learners were also assessed during the first term of the school year but teachers continued teaching following their respective Departmental programmes.

As such, performance snapshots were taken for different populations at particular points in time. For the:

(1) Control schools, two snapshots were available, one being a baseline;

(2) Phase 1 Treatment group, snapshots were available at four points in time, including a baseline, (over a 3 year period);

(3) Phase 2 Treatment group, three snapshots were available, including a baseline, (over a 2-year period); and

(4) Phase 3 Treatment group, two snapshots were available, one being a baseline.

An important design feature of the Project was the inclusion of “Control” schools that service the same area as the Jumpstart Programme schools. Although the Programme is largely an “intervention” rather than an academic research experiment, the ability to relate the results from the intervention to results in similar contexts without intervention strengthens the validity of whatever claims can be made from the Jumpstart Programme. This quasi-experimental design is commonly used for evaluating intervention programs where data has been collected from Treatment and Control, pre- and post-intervention, samples as against more robust randomised-control-trial (RCT) designs. The layout of the design is illustrated in Table 2 where $\bar{Y}$s are measured before and after the JumpStart intervention.

Table 2: The JumpStart quasi-experimental design

	Pre-intervention	Post-intervention
Treatment	$\bar{Y}_{Tr,Baseline}$	$\bar{Y}_{Tr,Endline}$
Control	$\bar{Y}_{C,Baseline}$	$\bar{Y}_{C,Endline}$

Our assumption is that the trend in the Foundation Phase mathematics performance in the Control group approximates what would have happened in the Treatment schools in the absence of the JumpStart intervention in Ekurhuleni East District during the period under evaluation. We confirmed that there was no similar or equivalent intervention in the Control schools during this period.

Research questions

Kelello collaborated with JumpStart to evaluate the impact that the intervention is making towards improving the teaching and learning of mathematics at the Foundation Phase in the sampled schools in the District of Ekurhuleni.

The focus of this report is on mathematics learning outcomes as measured using the EGMA tool. In the report we consider whether the JumpStart intervention has made any impact on learner attainment in this assessment at grade 1, 2 and 3 by considering whether:

- the EGMA attainment by grade differed between the Treatment and Control groups, at baseline and after one year; and
- there were changes evident over time (after 1 , 2 and 3 years of JumpStart intervention) in the cross-sectional performance by grade in Treatment schools.

Sample and sampling of schools

A two-stage sampling design was applied. In the first stage the Department of Education District drew the sample of schools in which they considered intervention would benefit both the school and the learners. At the sampled school random samples of FP learners are selected to take the EGMA test under the guidance of a trained JumpStart instructor. The method used to randomise the selection of learners is designed to avoid technical complexities associated with sampling but is powerful enough to ensure a random representation of learners. The method was aptly summarised as follows by one of the field workers:

> So, we arrive at a school and select a random number of learners in grade 3. We will have all the grade 3 line up from shortest to tallest. If there are 100 and we want 20 learners, we choose every 5th child. We select them and they come to the classroom. When they arrive, they get invigilated by an invigilator, and possibly 2 or 3 other intern helpers with the language issue.

The sampled learners are then grouped and accommodated in one classroom and each is given a tablet on which they complete the assessment as described by a field worker:

> They get a tablet which says Question 1 and they go through a few samples on how to respond. There is a teacher's guide that you can go through on how to do this and it invigilates quite carefully. The guidelines have prompts like: Please will everyone practice the following tests? Do you identify the number 3? Everyone clicks number 3, go around the class and make sure they have done it. Then they do the next one, please identify number 13 and if that is in the vernacular then it is in a different language. Just so all learners understand what this next test is going to do, we do the 10 questions where they are voiced on the projector: Please identify the following number? will you please compare the larger number or please finish the pattern? or here is a word problem please write down the answer.

It is important to emphasise that the Control schools were drawn from the same geographic and social context as the JumpStart Treatment schools. This meant that the population of learners who attend the two sets of schools should have similar characteristics and that any changes to levels of mathematics skills and knowledge in the Treatment schools could be ascribed to the JumpStart intervention. In the initial Control group as was identified by the District there were eleven schools whose quintile status ranged from Quintile 2 to Quintile 5 with three of the schools in Quintile 5 whereas all the Treatment schools were in Quintiles 1-3. Quintile 1 schools are the poorest (in terms resources such as school libraries, laboratories, school halls, sports facilities, etc) and generally rural schools while Quintile 5 schools are in the most affluent suburbs. Quintiles 1-3 are "no fee" schools and receive the greatest government subsidy. Quintile 4 and 5 schools receive the smallest subsidy but may charge parents school fees to augment their resources. As this system was introduced at the advent of democracy, it is now considered somewhat outdated and less reliable as an indicator of SES. Nevertheless, it offers some indication of historic SES.

Van den Berg et al (2019) refer to a 'dualistic education system in South Africa' and, drawing on Shepherd (2016), that Quintile 5 schools perform far better in the standardised reading assessment at Grade 5 (prePIRLS reading score) than Quintiles 1-4. Spaull and Kotze (2019) show the same bimodal pattern of learning outcomes for grade 3 mathematics, when distinguishing learners in Quintile 5 schools, from those in Quintiles 1-4:

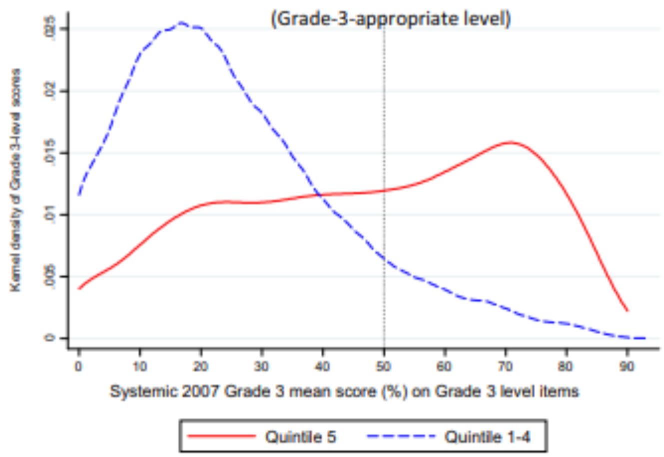

Fig. 1. Kernel density of mean Grade 3 performance on Grade 3 level items for Quintiles 1–4 (poorest 80% of students) and Quintile 5 (wealthiest 20% of students) (Systemic Evaluation, 2007).

Figure 1: Kernel density of mean grade 3 performance on grade 3 items for Q1-4

We therefore expect far higher performance in mathematics from quintile 5 schools. Given the quintile disparities among the Control and Treatment schools, one would expect higher attainment in mathematics in the Control schools in this sample because quintile level, as a proxy for SES, is known to be a strong predictor of attainment (Spaull, 2016). We, therefore, excluded the three Quintile 5 Control schools from the comparisons in the analysis.

The final sample of schools used in the analysis and reporting in the Treatment and Control groups, categorised by quintile status, has been presented in Table 3.

Table 3: Sample schools in the Jumpstart Programme

School quintile	Control schools	Treatment schools	Total
Quintile 1	0	1	1
Quintile 2	1	4	5
Quintile 3	3	14	17
Quintile 4	4	0	4
Quintile 5	0	0	0
	8	**19**	**27**

In the Treatment schools the EGMA tests were administered at the outset of engagement with JumpStart (baseline) and then one academic year thereafter. This gave a dataset where random samples of learners from the various cohorts were assessed at baseline, after 1, 2 and 3 years of JumpStart intervention. In the Control schools random samples of learners were assessed at baseline, and then one academic year thereafter.

EGMA data was obtained from 5 724 learners during the four years of the Programme as shown in Table 4.

Table 4: Numbers of assessed learners in the Jumpstart Programme

		Treatment schools		**Control schools**		**Total**	
	Baseline	**Year 1**	**Year 2**	**Year 3**	**Baseline**	**Year 1**	
Grade 1	500	438	316	146	358	176	1934
Grade 2	479	520	428	150	154	170	1901
Grade 3	441	516	436	149	172	175	1889
Total	1420	1474	1180	445	684	521	5724

We focused on attainment in Grade 1-3 where there was the most consistent data, and all group sizes exceeded 145 learners.

The EGMA test

The learner outcome data was drawn from the Core Early Grade Mathematics Assessment (EGMA), which is an orally administered assessment of the core mathematical competencies taught in early primary grades. EGMA development was undertaken by RTI International and funded by the United States Agency for International Development (USAID). The Core EGMA offers an opportunity to determine whether children are developing the fundamental skills upon which other mathematical skills build, and, if not, where efforts might be best directed (Platas, et al (2014). The Core EGMA is described as "an assessment of early mathematics learning, with an emphasis on number and operations" (Platas *et al*, 2014).

The EGMA test is a battery of seven testlets each focusing on assessing specific Number Sense skills. There are seven testlets in total each focusing on one set of mathematical skills:

- addition,
- subtraction,
- addition and subtraction together,
- number identification,
- number comparisons,
- number patterns (missing number) and
- word problems.

Given the specific focus on number, we refer to the underlying construct of the EGMA as "number sense".

In total there are 86 items that constitute the test. The test has been developed and used in different contexts with results that attest to its validity. The test had a reliability index of 0.96 and a Separation index of 5.12, which are considered to be well above the recommended minimum values of 0.80 and 2, respectively (Linacre, 2020). The Separation index is an indicator of the number of "ability groups" that the instrument detects from the testee responses. Any test that cannot detect at least two groups, competent and incompetent, among the testees is may not be useful for informing intervention decisions. The EGMA test has the ability to detect up to five ability groups among test respondents, which is phenomenal for a diagnostic instrument.

The Core EGMA has been designed for both summative and formative assessments. It has been used to (1) determine how students in a country are performing overall compared to its stated curriculum and (2) examine the effectiveness of specific curricula, interventions, or teacher training programs. In this case, the EGMA was used for the latter.

Platas *et al*. (2014) offer detailed explanation of the core EGMA's development, descriptions of its technical adequacy, as well as the details of the validity and reliability evidence collected are offered. Processes for local adaptation and training of assessors are also prescribed. The core EGMA is meant to be locally adapted to fit the needs of the local context, particularly with regard to language where it is administered in the language of learning and teaching for mathematics at Foundation Phase in the school. The EGMA test was translated into some African languages (at least Afrikaans and isiXhosa). The translations pertain to the instruction prompts, word problems and the naming of numbers.

EGMA test administration

The administration of the EGMA test is highly structured with protocols that attempt to establish a standardised process across institutions. The assessment is administered orally with an adult administrator reading out the instructions to a group of children. The children then make their choice of answer using a tablet. The administrator then moves on to the next question, reads this out and all children answer. This is the same way that Annual National Assessments were administered by teachers with their classes (DBE, 2012).

The script or guide was provided in English, and the adult administrator translated on the spot into the language of teaching and learning of the school. This is a design and data collection weakness as it means that the tests were not standardized and tightly scripted for each language. This places questions on the validity of data especially for the word problem questions. Test administration by interns

Part of the responsibility of the interns in the Jumpstart Programme is to assess learners as a way of monitoring the impact that the JumpStart initiative is making. Assessment involves administering quick but powerful short tests on learners using a well-researched and internationally recommended EGMA instrument. The EGMA tests are administered orally by trained interns and learners respond on tablets. Issues of language interference in learning mathematics are, to a large extent, appropriately mediated as children do not have to read but only listen to the instructor and respond on tablets.

The EGMA tests are administered in the language of teaching and learning (LOLT) in the Foundation Phase in each school. In the Ekurhuleni district the languages in which the tests were administer were English, isiXhosa, isiZulu, Sepedi and Sesotho. In a few schools more than one language was used for teaching learners from different language backgrounds. Necessary arrangements were made to make sure that the interns were competent the relevant languages. The strength of the EGMA tests is in being quick to administer and, therefore, quick to provide prompt feedback and intervention. The findings from the tests are used to monitor the efficacy the JumpStart interventions as they support the schools and the learners.

Data preparation

Preparation of data for analysis included merging all the separate files for each year cohort of learners in the JumpStart programme and compiling one Master database in Microsoft excel. Variables such as the quintile levels of the schools, number of teachers per school and school size were sourced from the DBE/GDE EMIS database and included into the Master database because they were considered to be important factors in intervening for improvement.

Data cleaning

Data cleaning involved removing from the dataset all learners who had zero responses and learners who were recorded as being in grades 4 and 5. Two Control schools that were identified as being in Quintile 5 were also removed from the dataset.

Adjustment of scores according to testing times

The majority of EGMA tests were written in April (04) of each academic year which runs from January to December in South Africa. However, in some years, and in some schools, EGMA tests were conducted later that April (with the latest being written in July). To adjust for these shifts in test administration times, we used the Control school baseline mean scores to calculate the expected ‘normal’ shifts in EGMA scores over one academic year (12 months). This data was drawn from the same 8 schools one year apart, and was stable from one year to the next by grade.

Table 5: Shifts in EGMA raw score and EGMA level

Control 1 year	Shifts in EGMA raw score	Shift in 'normal' EGMA level
Grade 1 to Grade 2	17,71 (≈2,6 SD)	0,48
Grade 2 to Grade 3	11,07 (≈1,3 SD)	0,47

1 month	Shift in EGMA raw score	Control 'normal' EGMA level
Grade 1 to Grade 2	1,48	0,04
Grade 2 to Grade 3	0,92	0,04

We could, therefore, use the expected "normal" shift over one academic year of 0,48 EGMA levels for grade 1s and 0,47 EGMA levels for Grade 2s. We calculated this as the expected shift (in raw score and EGMA level) over one month (considering an academic year to be 12 months). We used this to adjust the EGMA raw scores for students who wrote the test before or after April. To illustrate, a Grade 1 test score administered in February was adjusted upwards by 2 months (+2,98 EGMA raw scores and +0,08 EGMA levels)' and a Grade 1 test score administered in July was adjusted downwards by 3 months (-4,47 EGMA raw score and - 0,12 EGMA levels). In this way all of the EGMA results were adjusted to be as if they were all written in April.

Analysis

Data analysis included determining and comparing mean raw scores obtained by learners in the EGMA test and the mean EGMA levels attained by learners. We also determined frequencies to determine proportions of learners who function at various EGMA Levels. Largely, therefore, we used descriptive statistics. We used t-test statistics to compare mean score differences, where they existed, for statistical significance. To determine the impact of the intervention we calculated Cohen's D which measures the distance between two population means and has been used in many studies to calculate "effect sizes" (ES). We employed the use of Cohen's D co-efficient to estimate the impact of the Jumpstart intervention on Foundation Phase learners in Ekurhuleni District. Cohen's D measures the distance between two means from two populations. Although there are divergent views on "acceptable" "effect size", Rhea (1988) gives a guiding rule that, values less than 0.41 indicate "small" effect, values of 0.41-0.70 indicate "moderate" effect and values greater than 0.70 show "large" effect.

Preliminary analysis, using modern item response theory (IRT) models, indicated that there were many items on which no score was allocated, because many learners were not able to complete the timed sections of the test within the prescribed time. Information on whether learners were not able to complete the test because of time constraints or they actually were not able to respond to the items is important and will be interrogated in the second stage of the analysis when learner-level details are investigated. A factor that may not be overlooked is early onset of fatigue, especially among younger children. Onset of fatigue may make it difficult to know whether lack of responses in a timed test is due to lack of knowledge or is just that children got tired. For instance, items that were not attempted at all by learners made the test to appear more difficult than it was. A further analysis will seek to interrogate these possibilities in order to increase the efficacy of the testing process.

Results

The main focus of the report is on establishing whether the JumpStart intervention made any difference in the Treatment schools in terms of Foundation Phase learners' competencies in mathematics and, if so, whether sustaining the intervention also sustained the outcomes. The overall presentation of the results covers findings on the *status quo* in both the Treatment and Control schools before the intervention and then includes changes that were observed within one year of the intervention. We then used the cross-sectional data to track changes that occurred in learners' mathematics competencies as cohorts of learners in the Treatment schools progressed from grade to grade over the four years (including baselines) of intervention. We compared learner attainment in the Treatment schools with their counterparts in the Control schools and with their own baseline attainment.

We report the results using descriptive statistics applied to the raw scores and to EGMA Levels, mainly to enable users of the results to have a wider choice on the format of reporting that best suits their needs. For instance, while the use of raw scores for reporting may be a common practice among teachers, the use of the innovative EGMA Levels provides a unique opportunity for quick screening of learners who function at various levels and may be more ideal than raw scores for informing relevant interventions. This is made possible by the special configuration of EGMA mathematics competencies in a hierarchical order of cognitive demand. Finally, we use the technique of *difference-in-differences* to estimate the overall effect of the JumpStart intervention programme in the Treatment schools and compare this effect with what has been reported in the literature on intervention project evaluations.

Comparability of Treatment and Control schools

To verify the comparability of Control and Treatment school performance in this Programme, we compared the grade 1 EGMA results in the Treatment schools with those of their counterparts in the Control schools in terms of both raw scores and EGMA Levels. The mean baseline raw scores and EGMA Levels (and their respective standard deviations) for the Treatment and Control schools at grade 1 level were calculated and have been presented in Table 5.

Table 6: Grade 1 baseline scores and EGMA Levels

Treatment (*n*=500)		**Control (*n*=358)**	
Mean raw score (SD)	Mean EGMA Level (SD)	Mean raw score (SD)	Mean EGMA Level (SD)
18,46 (10,34)	1,38 (0,48)	21,85 (6,81)	1,51 (0,50)

A marked observation that we made was that learners in the control schools were beginning from a higher mathematics competency baseline than their counterparts in the Treatment schools when the EGMA tests were administered at the end of the first term. The mean differences, in both raw scores and EGMA Levels, were statistically significant at the p=.05 level. We know that weaker foundations in mathematics tend to widen over time. Spaull and Kotze (2019) found that, in relation to mathematics learning:

> by Grade 3, children in Quintiles 1–3 are already three years' worth of learning behind their Quintile 5 peers and that this gap grows as they progress through school to the extent that by Grade 9 they are four years' worth of learning behind their Quintile 5 peers. (p.23)

In the absence of the intervention, it would, therefore, be expected that the Control schools (with higher proportions of quintile 4 schools) would start with stronger EGMA results that the Treatment schools drawn from quintiles 1-3 and that these differences would grow wider over time.

Baseline EGMA results

As was indicated in the section on the research design, each phase of the project started with establishing a baseline and trends were then followed from year to year over three years. The baseline EGMA results in the Treatment and Control schools, consolidated over three years, have been summarised in Table 6.

Table 7: Baseline EGMA results consolidated over three year

Treatment schools			Control schools		
	Mean score (SD)	Mean EGMA Level (SD)		Mean score (SD)	Mean EGMA Level (SD)
Grade 1 (*n*=500)	18.46 (10.34)	1.38 (0.48)	Grade 1 (*n*=358)	21.85 (6.81)	1.51 (0.50)
Grade 2 (*n*=479)	33.11 (11.06)	1.81 (0.50)	Grade 2 (*n*=154)	39.09 (8.65)	1.99 (0.50)
Grade 3 (*n*=441)	42.86 (12.18)	2.16 (0.75)	Grade 3 (*n*=172)	50.03 (7.75)	2.61 (0.80)

For all the three grades, both the mean raw scores and the mean EGMA Levels tended to be higher for the control than treatment schools - an observation which again can be ascribed to the fact that there was an over-representation of quintile 4 in the Control group while all the treatment schools were in the quintiles 1-3 schools characterised by a relatively low SES. Not only were the mean raw scores lower in treatment schools, but, as can be observed from the relatively higher standard deviations (SDs), there seemed to be higher variability in the distribution of learner scores in treatment than control schools. But what was also a noteworthy, though to be expected, observation was the fact that performance in the EGMA test across the grades in the same group (treatment or control) was lowest at grade 1, higher at grade 2 and the highest at grade 3 level.

Emerging trends in EGMA results over time

To determine whether participation in the Jumpstart intervention made a difference in learner mathematics competency, a trend analysis was conducted over a period of four school years (including baseline) from the commencement of the intervention. We used the repeated cross-sectional data to report EGMA attainment trends over time in the treatment schools and then we compared trends between treatment and control schools over one school year to determine the impact that can be ascribed to the JumpStart intervention in the treatment schools.

In Figure 1 an illustration of the trends over four years (including baseline) in the treatment schools is given. Evidently, the overall trend was upward, slightly flat at grade 1 as would be expected since grade 1 has new admissions every year, but definitely steeper for grades 2 and 3 with prolonged exposure to the programme.

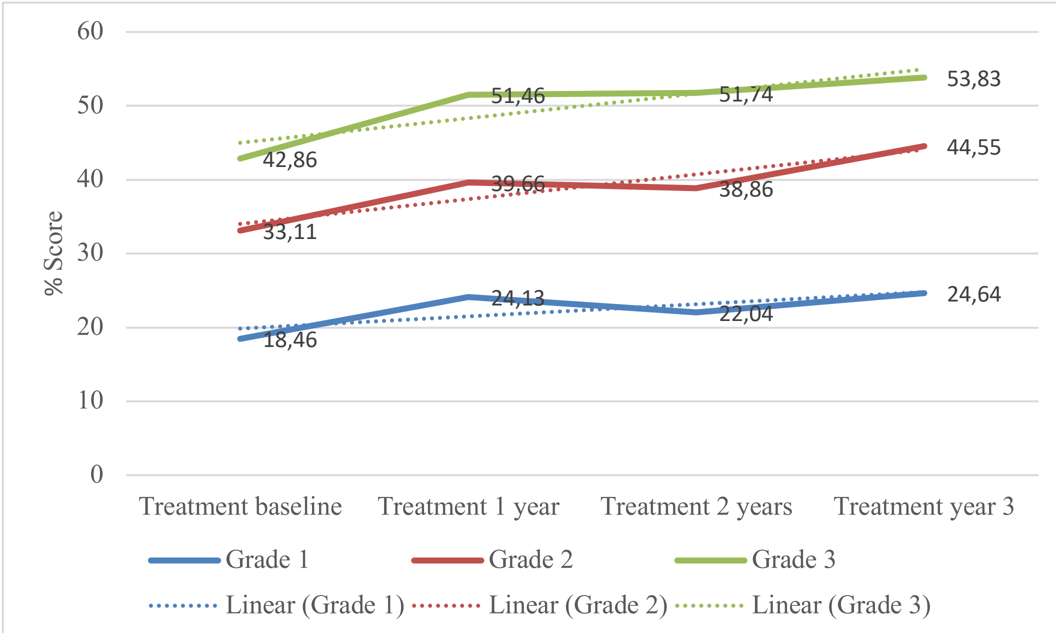

Figure 2: Mean EGMA attainment in Treatment schools over three school years

The pattern in the control schools over one year is shown in Figure 2. There were hardly any changes in the control schools in the period of one school year under consideration, a pattern which would be expected in the Treatment schools as well, if there was no intervention.

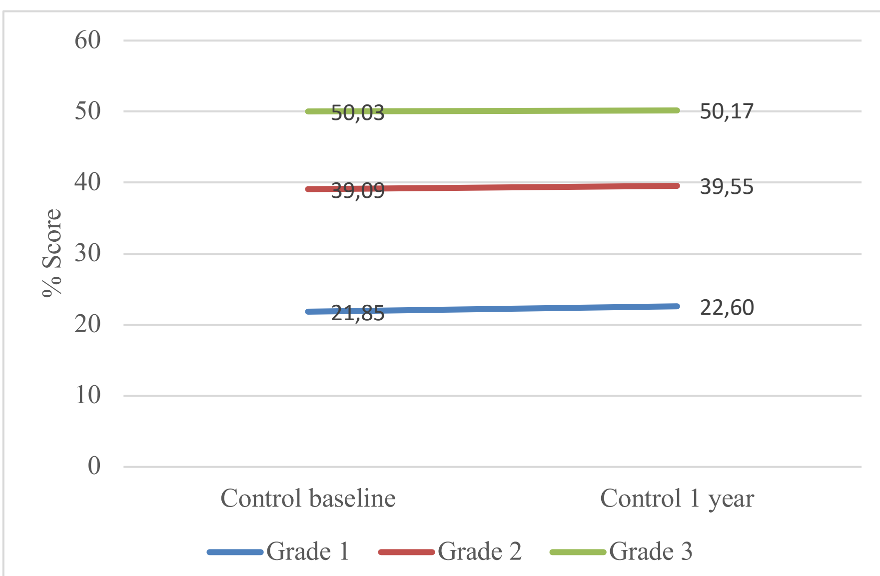

Figure 3: Mean EGMA attainment in Control schools over one school year

In Table 7 the changes in attainment scores in treatment and control schools over one school year are compared, expressed in terms of the number of baseline standard deviations, i.e. we answer the question: By how many standard deviations did the mean score change in one school year? For the Treatment schools there were increases ranging from approximately +0.6 to +0.7 of the baseline standard deviations across grades 1 to 3 whereas in the control schools the increases diminished from nearly +0.1 to +0.02 of a standard deviation across the same grades.

Table 8: Patterns of change in EGMA results in Treatment and Control schools over one year

Treatment			Control		
	Mean baseline (SD)	**1 year (Change in SDs)**		**Mean baseline (SD)**	**1 year (Change in SDs)**
Grade 1 (*n*=500; 438)	18.46 (10.34)	24.13 (+0,55 SD)	Grade 1 (*n*=358; 176)	21.85 (6.81)	22.60 (+0,11 SD)
Grade 2 (*n*=479; 520)	33.11 (11.06)	39.66 (+0,59 SD)	Grade 2 (*n*=154; 170)	39.09 (8.65)	39.55 (+0,05 SD)
Grade 3 (*n*=441; 516)	42.86 (12.18)	51.46 (+0,71 SD)	Grade 3 (*n*=172; 175)	50.03 (7.75)	50.17 (+0,02 SD)

A cross-sectional analysis of learner cohort attainment

From analysis of cross-sectional data that spanned the four years of the JumpStart intervention we followed the attainment of FP cohorts of learners in mathematics in the Treatment schools in terms of changes in EGMA Levels over time as shown in Table 8. We compared this attainment to what happened in the Control schools as shown in Table 9. A similar comparative tracking of Treatment and Control cohorts was also done using raw scores (Table 10 and Table 11, respectively).

Table 9: Cross-sectional cohort attainment in Treatment schools over time in EGMA Levels

	Treatment baseline mean (SD)	**Treatment 1 year mean (SD)**	**Treatment 2 years mean (SD)**	**Treatment 3 years mean (SD)**
Grade 1 (*n*=500; 438; 268; 146)	1.38 (0.48)	1.45 (0.51)	1.40 (0.74)	1.51 (0.50)
Grade 2 (*n*=479; 520; 395; 150)	1.81 (0.50)	2.00 (0.55)	2.01 (0.74)	2.07 (0.45)
Grade 3 (*n*=441; 516; 406; 149)	2,16 (0.75)	2,79 (0.96)	2.71 (0.82)	2,67 (0.79)

Table 10: Cross-sectional cohort attainment in Control schools over time in EGMA Levels

	Control baseline mean (SD)	**Control 1 year (SD)**
Grade 1 (*n*=176; 176)	1.51 (0.50)	1.39 (0.52)
Grade 2 (*n*=154; 170)	1.99 (0.50)	1.98 (0.47)
Grade 3 (*n*=172; 175)	2.61 (0.80)	2.46 (0.79)

Two important observations can be made from tracking the cohorts of learners. Firstly, the first grade 1 cohort in the treatment schools (M=1.38; SD=0.48) started below the grade 1 cohort EGMA Level in the control schools (M=1.51; SD = 0.50). When these treatment learners were in Grade 3 (M=2.71; SD=0.82), they had caught up with the grade 3 learners in the control schools (M=2.61; SD=0.80), and actually surpassed them by +0.10 EGMA Level which is approximately 0.20 of a baseline SD.

But compared to their own baseline, after three years in the JumpStart intervention (including baseline), this cohort had improved grade 3 performance by 0.73 SD ((2,71-2.16/0.75)).

Secondly, the second grade 1 cohort in the treatment schools (M=1.45; SD=0.51) also started above the grade 1 cohort in the control schools (M=1.51; SD=0.50). Although this cohort's attainment at grade 3 (M=2,67; SD=0,79) was lower than their previous counterparts (M=2,71; SD=0.82), these Treatment learners had also caught up with the grade 3 learners in the control schools (M=2.61; SD=0.80) and had actually surpassed them by an average of +0.06 EGMA Level which is approximately 0.13 of a baseline SD. Again, this cohort did not only improve in relation to their counterparts in the control schools, but they also improved from their own baseline by (2.67 - 2.16)/0.75 which is approximately 0.68 SD.

Similar changes were observed when raw scores were used to track the two cohorts as shown in Table 10 for treatment schools and Table 11 for control schools and improvements of approximately 0.16 SD and 0.35 SD, respectively, were realised when compared to the Control school cohorts. When comparing the treatment schools with their own baselines, using raw scores, the improvements in performance for the two cohorts were 0.73 SD and 0.90 SD, respectively.

Table 11: Cross-sectional cohort attainment in Treatment schools over time in raw scores

	Treatment baseline mean (SD)	**Treatment 1 year mean (SD)**	**Treatment 2 years mean (SD)**	**Treatment 3 years mean (SD)**
Grade 1 (*n*=500; 438; 316; 146)	18.46 (10.34)	24.13 (7.44)	22.04 (6.97)	24.64 (7.44)
Grade 2 (*n*=479; 520; 428; 150)	33.11 (11.06)	39.66 (8.59)	38.86 (8.59)	44.55 (8.95)
Grade 3 (*n*=441; 516; 436; 149)	42.86 (12.18)	51.46 (8.76)	51.74 (10.51)	53.83 (10.51)

This can be compared to the cross-sectional data for the control schools which gives EGMA raw score per grade, as shown in Table 12.

Table 12: Cross-sectional cohort attainment in Control schools over time in raw scores

	Control baseline mean (SD)	**Control 1 year (SD)**
Grade 1 (*n*=358; 176)	21.85 (6.81)	22.60 (6.97)
Grade 2 (*n*=154; 170)	39.09 (8.65)	39.55 (8.59)
Grade 3 (*n*=172; 175)	50.03 (7.75)	50.17 (10.51)

In general, there was an upward trajectory in performance whereby every cohort in the Jumpstart intervention out-performed their predecessors at the same grade.

We used the repeated cross-sectional data to reflect on EGMA attainment trends over time in the control and treatment schools. (Figure 1) to reflect on changes evident over time (after 1, 2 and 3 academic years of JumpStart intervention) in the cross-sectional performance by grade in treatment schools

.

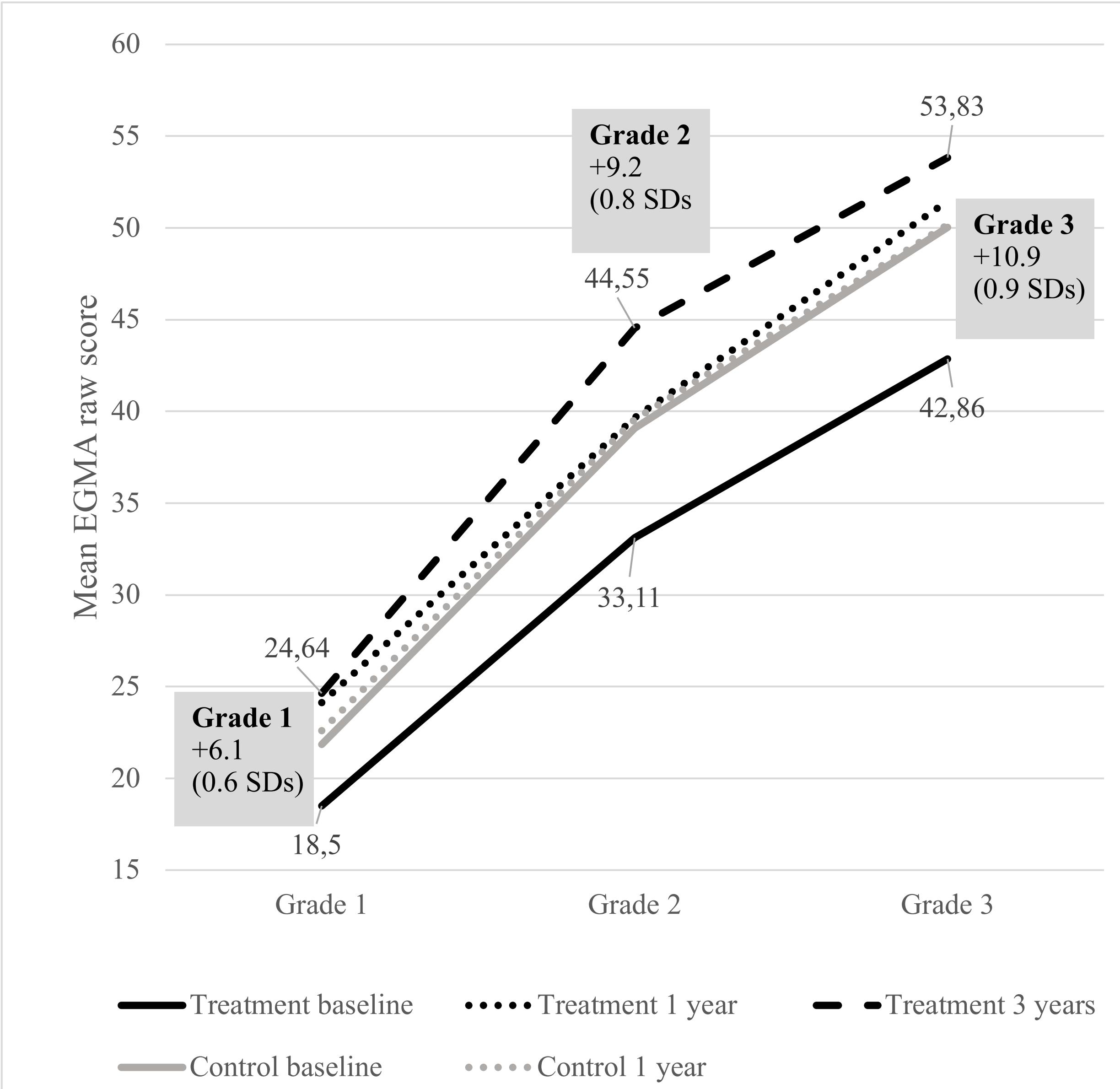

Figure 1: EGMA mean raw scores for treatment and control schools

There were clear improvements in the EGMA mean attainment as JumpStart was embedded in each school, and learners had the benefit of multiple years of intervention. The highest mean attainment by grade level was evident after 3 years of JumpStart intervention.

Shifts in attainment across grades

We investigated changes or shifts that occurred as cohorts of learners progressed from one grade to the next both in treatment and control schools, in terms of proportions of learners who function at various EGMA Levels, the statistical significance of mean attainment measures and also in terms of *difference-in-differences (DiD* or Delta) measures.

Proportions of learners functioning at various EGMA Levels

In Figure 3 the baseline proportions of learners who were functioning at various EGMA Levels are shown for grades 2 and 3. For each grade the proportion of learners at the two highest EGMA Levels (L3 and L4) was higher in the control than in the treatment schools. However, after one year of the JumpStart intervention, there was not only a closing of the gaps for each grade, but there was a total reversal of the observed phenomenon as shown in Figure 4. For each of grades 2 and 3 the proportion of learners functioning at the two highest EGMA Levels (L3 and L4) was markedly higher in the treatment than control schools even though the latter were predominantly Quintile 4 schools while the former were no-fee schools in Quintiles 1-3.

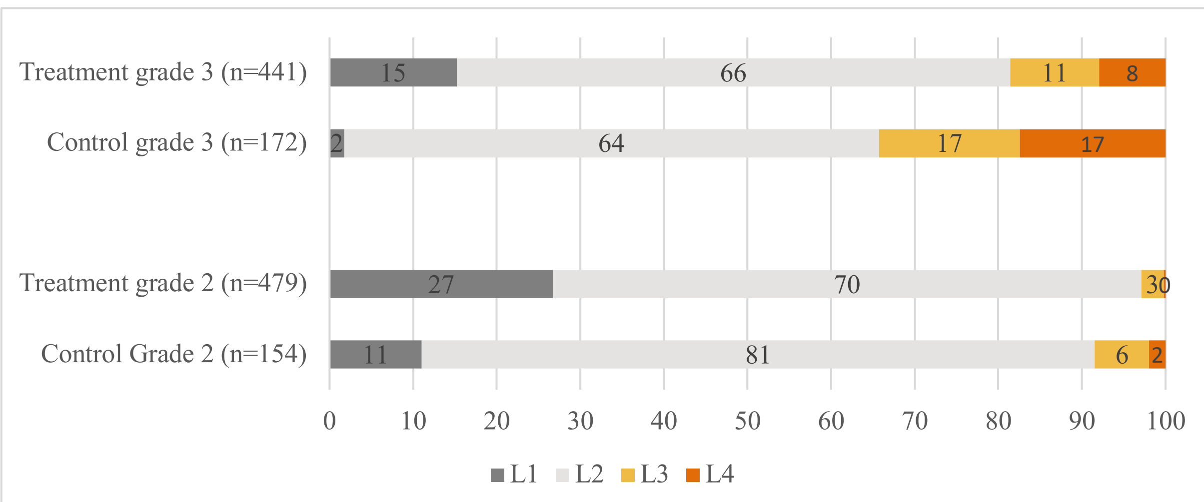

Figure 4: Baseline proportion of learners at different EGMA Levels

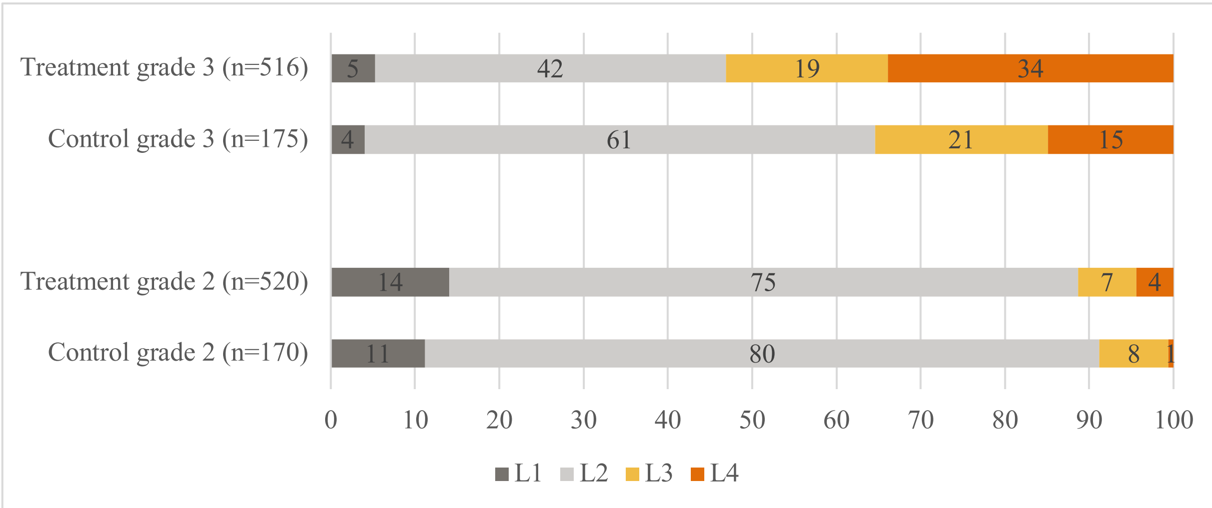

Figure 5: Proportion of learners at different EGMA Levels 1 year after intervention

The JumpStart intervention in Foundation Phase mathematics performance brought about a complete reversal in the known pattern of performance whereby schools in Quintiles 4 and 5 tend to outperform the no-fee schools in Quintiles 1-3 and thus the intervention serves to bridge the equity gap.

Statistical significance of attainment measures

We tested the changes in mean EGMA Levels for statistical significance as cohorts of learners in the treatment schools progressed from one grade to the next. T-tests showed that there were statistically significant improvements in mean EGMA Levels as learners in the intervention schools progressed from grade 1 baseline (M=1.38; SD = 0.48) to grade 2 (M=2.00; SD=0.62), t(972) = 17,73, p=0.00. Similarly, there were statistically significant improvements in mean EGMA Levels as learners progressed from grade 2 (M=1.81; SD=0.50) to grade 3 (M=2.79; SD=0.96), t(784) = 20.39, p=0.00.

Intervention effects in terms of Cohen's D

In addition to tracking trends in attainment for the cohorts over time we also investigated shifts in EGMA Levels as learners progressed from grade to grade and compared these between Treatment and Control schools. A "shift" in this context refers to the growth (Cohen's D or delta) in EGMA

Level which denotes an increase in the mathematics knowledge of the learner as measured by the EGMA Level attained.

The shifts in EGMA Levels as learners progressed from grade 1 to grade 2 (Grade 1-2) and from grade 2 to 3 (Grade 2-3) for Treatment and Control schools are shown in Figure 3. Treating the shifts in the Control schools as the "normal", i.e. expected growth in the absence of the intervention (also known as the counterfactual), we note that in the Treatment schools the mean EGMA Levels improved by Delta=0.63 – 0.48 = 0.15 which translates to nearly 0.30 of baseline standard division (M=1.51; SD=0.50; n=176) more mean EGMA Level for the grade 1-2 shift.

For the grade 2-3 shift the improvement was Delta =0.93 - 0.47 = 0.46 which amounts to approximately 0.92 of baseline standard deviation (M=1.99; SD=0.50; n = 154) more mean EGMA Level than "normal".

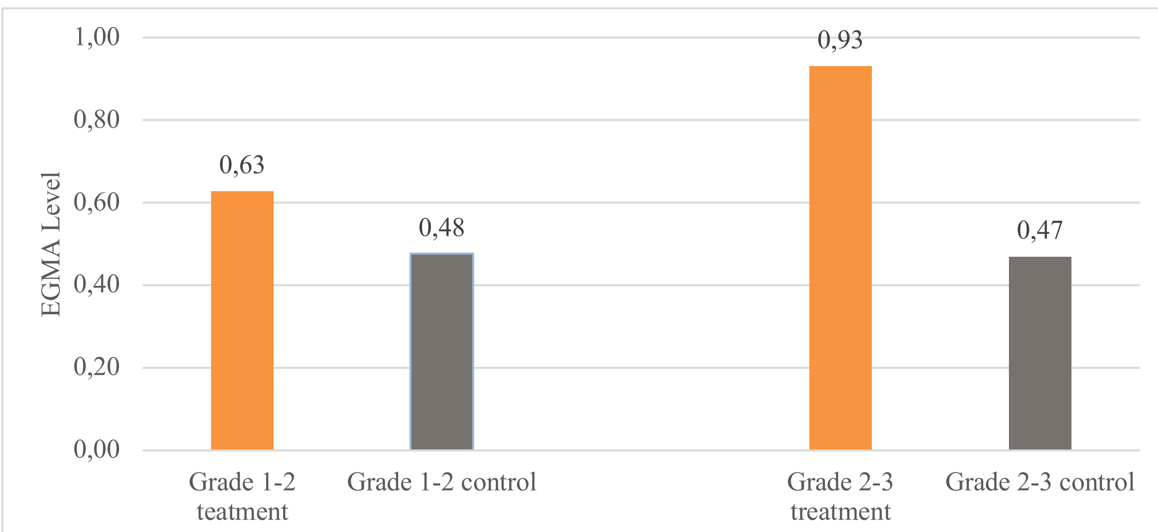

Figure 6: Comparing shifts in EGMA Levels between Treatment and Control schools

Estimating the overall *'difference-in-differences'* programme effect

The quasi-experimental design made it possible to use '*difference-in-differences*' (*DiD*) analysis, a technique commonly used for evaluating intervention programs where data has been collected from treatment and control, pre- and post-intervention, samples which are less robust than randomised-control-trial (RCT) designs.

We ran a regression analysis of the learner post-intervention test scores (Y) on three dummy variables, viz a variable (D_1) that denoted whether the learner was in the treatment or control sample, the second (D_2) which reflected whether the learner attained a post-intervention score or not and a third (D_1*D_2) which denoted the interaction of the first two dummy variables. For D_1 each learner was allocated a score of 1 if they were in the treatment group and 0 if they were in the control group, for D_2 a learner was allocated a 1 if they obtained a post-intervention score or 0 if they did not and the third dummy (D_1*D_2) was a time and treatment interaction of the two obtained by multiplying D_1 and D_2. For instance, for a learner who was in the treatment group, $D_1=1$, $D_2=1$ and $D_1*D_2=1$ whereas for a learner who was in the control group, $D_1=0$, $D_2=1$ and $D_1*D_2=0$ while for a learner who was in the treatment group but did not have a post-intervention score the dummy variables were $D_1=1$, $D_2=0$ and $D_1*D_2=0$.

Our regression model for estimating the *DiD* effects was:

$$\hat{Y} = \beta_o + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_1 * D_2 + (\beta_4 \mathbf{X} + \epsilon) \quad (1)$$

where the dependent variable $\hat{Y}$ is the predicted post-intervention test score, β_o is a constant (intercept) in the absence of any intervention and $\beta_{i=1,2,3}$ are coefficients of the dummy and interaction variables, $\beta_4 X$ are co-variables that might confound results in the control group but can be controlled for and ϵ is an error term associated with the measurement. Because both treatment and control schools were in the same district, we assumed that both the co-variables $\beta_4 X$ and the error term would be negligibly small. With these caveats, we ran a regression that resulted in the summary output in Table 12.

Table 13: Regression summary output for grade 2&3 EGMA levels

SUMMARY OUTPUT (Grade 2&3 EGMA Levels)						
Regression Statistics						
Multiple R	0.29					
R Square	0.08					
Adjusted R Square	0.08					
Standard Error	0.77					
Observations	1542					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	3	83.29	27.76	47.26	0.00	
Residual	1538	903.53	0.59			
Total	1541	986.82				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
β_o(Intercept)	2.33	0.05	44.82	0.00	2.23	2.43
β_1(Treatment/Control)	-0.41	0.06	-6.38	0.00	-0.53	-0.28
β_2(Post-intervention score)	-0.11	0.07	-1.59	0.11	-0.24	0.02
β_3(Time/treatment Interaction)	0.69	0.08	8.30	0.00	0.52	0.85

The mean '*difference-in-difference*' (*DiD*) for grade 2&3 EGMA levels, represented by β_3, the coefficient of the time/treatment interaction, was 0.69 ≈ 0.7 of an EGMA Level and was statistically significant (p=0.00) with a 95% Confidence Interval (CI) [0.52,0.85]. The baseline combined mean EGMA Level for the treatment group grades 2&3 was 1,93 ≈ 1,9 (n = 441, SD=0.77). The *DiD* or *effect size* was, therefore, an improvement equivalent of 0.9 SD (0.7/0.77 ≈ 0.9) of an EGMA Level over a period of four years (including baselines) of intervention.

Average "effect sizes" of magnitude ranging from 0.35 to 0.61 within one year of intervention were reported by Fleisch et al (2016) in an intervention for improving literacy and mathematics in Gauteng schools and were considered to be on the "large" side. For JumpStart, after an intervention spanning three school years, the effect had increased from 0.30SD after two years to approximately 0.93SD. This is much higher than, for instance, interventions that included general pedagogical interventions which improve teaching quality, as reported by Conn (2017) in his meta-analysis, and much higher than effect sizes reported by Snilsveit et al. (2015) in an intervention that focused on remedial education in mathematics.

Conclusion

The JumpStart intervention in effective teaching of mathematics at the Foundation Phase in the Education District of Ekurhuleni, in Gauteng, has produced results that need to be followed in order to address the low performance and disparities in the teaching of mathematics at the Foundation Phase in South Africa. Kelello, in collaboration with JumpStart, evaluated the intervention to determine the extent to which learners who participated in the Project developed and improved their 'number sense'. The evaluation took the form of a desktop exercise whereby data collected by JumpStart interns from the schools that they supported was analysed to determine levels of "number sense" attained by learners in two of three academic periods of the intervention.

In terms of the research questions that framed the approach to this evaluation, the following were the key findings presented in the report:

The evaluation findings show that, the JumpStart intervention made a large impact, as measured through "effect size", on the target learners in terms of improving their competencies in mathematics. Learners in the Jumpstart Programme schools demonstrated markedly higher levels of 'number sense' as measured through an internationally standardised Early Grade Mathematics Assessment (EGMA) instrument. Not only did the mathematics competencies of the learners in the intervention improve convincingly, but the learners in the intervention outperformed their counterparts in schools that operate in better off socio-economic circumstances as measured through quintile as a proxy for Socio-economic Status (SES). The influence of SES on learner attainment is an issue of national concern in South Africa. Therefore, an intervention that has the potential to narrow or close the equity gap associated with poverty should be prioritised for development as part of national initiatives to give every child an opportunity to learn, particularly in the essential subject such as mathematics.

The findings from this evaluation corroborate those made by Roberts (2020) earlier when the JumpStart intervention involved only 10 schools. At the time of this evaluation the number of schools in the intervention program had almost doubled but the outcomes continued to improve. This suggests that the intervention does not only have the efficacy to improve the teaching and learning of mathematics, but it has the potential to make the same if not a greater impact even when taken to scale.

The most noteworthy observation made in the evaluation were the gains made by participating learners in increasing their portfolios of mathematical knowledge and skills within relatively short time frames of exposure to the JumpStart program. We noted in the evaluation that, although grade 1 learners start school with serious deficits in the knowledge and skills expected, learners in the first year of the Foundation Phase benefitted the most from this project in terms of increased 'number sense'. But all learners experienced higher rates of increase in their mathematical knowledge and skills when exposed to sustained intervention as was the case in the Jumpstart Project which spanned four years at the time of this evaluation. Each year of participation in the intervention helps the learner to either catch up with others or exceed expectations.

Key recommendations that arise from this evaluation include:

- That prospects of securing funding to expand the Project to more schools in the same area, at first, but beyond the Gauteng borders in the long run. With the potency that has been demonstrated in this evaluation report, the intervention has the potential to be taken to scale.
- More research needs to be conducted into finer nuances of how the intervention, through the use of interns, manages or deals with issues of pedagogy as part of the entire teachers' professional skill, whether the issue of language of teaching and learning does become a possible barrier or the semi-independent use of tablets is adequate to address this concern.
- That this evaluation is extended to include a focus on qualitative dimensions that includes the evaluation of the actual intervention at classroom level be commissioned. This evaluation focused on the achievement of learning outcomes by learners without the benefit of evaluating the inputs that are antecedent to the outcomes. An elaborate evaluation should include evaluating how interns mediate learning in the classrooms, the extent to which learners complete ensuing assignments after EGMA results are obtained and whether there is synergy between what JumpStart interns and permanent school teachers do in the same classes. This information will position the project to maximise its impact.

References

Brombacher and Associates (2019). *The NumberSense Mathematics Programme,* https://www.numbersense.co.za/about/about-the-workbooks, accessed on 9 March 2019

Brombacher and Associates (2019). *Making the most of NumberSense workbooks*https://www.numbersense.co.za/assets/images/file_download/Workshop_Making-the-most-of-the-NSWKBKs.pdf, accessed on 19 August 2019

Conn, K. M. (2017). Identifying effective education interventions in Sub-Saharan Africa: A meta-analysis of impact evaluations. *Review of Educational Research*, 87(5), 863– 898.

Department of Basic Education. (2015). *Education Districts, education District Profiles* 2015, https://www.education.gov.za/Informationfor/EducationDistricts.aspx, accessed on 4 April 2019, Government of South Africa: Pretoria.

Department of Basic Education. (2012). *Annual National Assessments 2011 Report.* Pretoria: Department of Education.

Department of Basic Education. (2015). *Annual National Assessments 2014 Report.* Pretoria: Department of Education.

Fleisch, B., Schöer, V., Roberts, G., & Thornton, A. (2016). System-wide improvement of early-grade mathematics: New evidence from the Gauteng Primary Language and Mathematics Strategy. International Journal of Educational Development, 49, 157– 174.

Linacre, J. M. (2020). Winsteps®Rasch measurement computer program User's Gude. Beaverton, Oregon,: Winsteps.com

Platas, L.M., Ketterlin-Gellar L., Brombacher A., and Sitabkhan Y. (2014). *Early Grade Mathematics Assessment (EGMA) Toolkit.* RTI: Research Triangle park, North Carolina. 1-49

Rhea, M. R. (2004). Determining the magnitude of treatment effects in strength training research through the use of the effect size. *Journal of strength and conditioning research*, *18*, 918-920.

Roberts (2020 in press) Assessing early grade mathematics learner outcomes using m-learning: The case of JumpStart South Africa in Auer, M.E. (Ed) *Internet of Things, Infrastructures and Mobile Applications*, Springer.

SACMEQ, (2017). The SACMEQ IV Project in South Africa: A Study of the Conditions of Schooling and the Quality of Education. http://www.sacmeq.org/sites/default/files/sacmeq/reports/sacmeq-iv/national-reports/sacmeq_iv_project_in_south_africa_report.pdf Accessed on 12 July 2020.

Shepherd, D. (2016). What is the limit? Performance ceilings at secondary school and their relation to socioeconomic status. Unpublished paper presented at ReSEP workshop, 21 November 2016. Stellenbosch. https://doi.org/10.1016/j.ijedudev.2015.01.002

Spaull, N. (2016). Poverty & privilege: Primary school inequality in South Africa. *International Journal of Educational Development*, *33*(5), 436–447. https://doi.org/10.1016/j.ijedudev.2012.09.009

Spaull, N. & Kotze (2019). Starting behind and staying behind in South Africa The case of insurmountable learning deficits in mathematics, *International Journal of Educational Development* 41 (2015) 13–24