SA-TIED
Southern Africa – Towards Inclusive Economic Development

SA-TIED YOUNG SCHOLARS PROGRAMME

Words can change worlds

An impact evaluation of Shine Literacy

Charlotte-Kathrin Stollberg

SA-TIED Working Paper #16 | May 2018

UNITED NATIONS UNIVERSITY UNU-WIDER
national treasury Department: National Treasury REPUBLIC OF SOUTH AFRICA
INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE IFPRI
planning, monitoring & evaluation Department: Planning, Monitoring and Evaluation REPUBLIC OF SOUTH AFRICA
the dti Department: Trade and Industry REPUBLIC OF SOUTH AFRICA
SARS
TIPS

Young Scholars

This paper was produced as a part of the SA-TIED Young Scholars' programme. The programme is a part of SA-TIED's capacity building initiatives, designed to support the development of skills and capabilities of university students interested in public sector research. Every year, SA-TIED engages Masters' level students from Southern Africa to become Young Scholars.

SA-TIED Young Scholars work with top academics and officials in their research fields to complete original research projects as a part of the research component of their Masters' degree.

About the programme

Southern Africa –Towards Inclusive Economic Development (SA-TIED)

SA-TIED is a unique collaboration between local and international research institutes and the government of South Africa. Its primary goal is to improve the interface between research and policy by producing cutting-edge research for inclusive growth and economic transformation in the southern African region. It is hoped that the SA-TIED programme will lead to greater institutional and individual capacities, improve database management and data analysis, and provide research outputs that assist in the formulation of evidence-based economic policy.

The collaboration is between the United Nations University World Institute for Development Economics Research (UNU-WIDER), the National Treasury of South Africa, the International Food Policy Research Institute (IFPRI), the Department of Monitoring, Planning, and Evaluation, the Department of Trade and Industry, South African Revenue Services, Trade and Industrial Policy Strategies, and other universities and institutes. It is funded by the National Treasury of South Africa, the Department of Trade and Industry of South Africa, the Delegation of the European Union to South Africa, IFPRI, and UNU-WIDER through the Institute's contributions from Finland, Sweden, and the United Kingdom to its research programme.

Southern Africa - Towards Inclusive Economic Development

SA-TIED | Working Paper 16 | May 2018

Words can change worlds

An impact evaluation of Shine Literacy

Charlotte-Kathrin Stollberg

ABSTRACT

The recent PIRLS report painted a dire picture of South Africa's literacy: 80% of Grade 4 learners are unable to read for meaning. With a large body of research confirming how reading literacy holds predictive validity for later child development and academic success, Shine Literacy offers an intervention that is set as a measure for those who struggle to read. This quasi-experimental impact evaluation estimated the treatment effect of Shine Literacy via difference-in-differences. Average treatment effects range between 0.6 to 1.9 standard deviations. IsiXhosa and 'At risk' learners are the main beneficiaries of the programme.

JEL classifications: D04, H43, I24

Keywords: education, early literacy, difference-in-differences, programme evaluation, South Africa

1 INTRODUCTION: THE SOUTH AFRICAN LITERACY CRISIS

Reading literacy is a key problem among South African learners. The latest results from the 2016 Progress in International Reading Literacy Study (PIRLS) on Grade 4 reading achievements across 50 participating developed and developing countries indicated that eight out of ten Grade 4 learners in South Africa are unable to read for meaning (Mullis et al. 2017). Learners cannot locate clearly listed information in text or make inferences about stated events or actions (Mullis et al. 2017). The 2011 PIRLS, which considered South African Grade 5 learners, showed that even in higher grades the picture is dismal: 29% of those students did not reach the pre-PIRLS international benchmark for literacy (Pretorius and Spaull 2016). Similarly, the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) reports have indicated that 27% of South African Grade 6 learners are functionally illiterate,[1] with only very little improvement between 2000 and 2007 (Moloi and Chetty 2011). With fewer than 40 correct words per minute, learners are essentially rendered as 'non-readers' (Pretorius and Spaull 2016: 1456).

The results from the SACMEQ and PIRLS reports need to be seen in the context of the extensive public spending on education in South Africa, which has been 5–6% of GDP since 1994 (World Bank 2016). This is relatively high, internationally speaking (World Bank 2016). Given that about 99% of learners are enrolled, school enrolment itself does not seem to be the issue (World Bank 2016). Instead, South Africa's education crisis can be linked to the lack of quality inputs, not access – a phenomenon encountered repeatedly in developing countries (Pretorius and Spaull 2016). As a consequence of a poorly performing education sector, the proportion of unemployable low-skilled youth is increasing, ultimately leading to a widening of the already stark income and wealth inequality, and to a further halt on the move towards a knowledge-based South African economy (McCarthy and Oliphant 2013).

Universally, as one of the primary skills acquired at school, reading literacy is defined not only as the ability to understand, use and reflect on text to develop reading knowledge and skills, but also, ultimately, to participate actively and successfully in society (Howie et al. 2008; Linnakylä et al. 2004; Pretorius and Spaull 2016). Illiterate individuals are at risk of being excluded from working and studying, but also from cultural or social life (Linnakylä et al. 2004). To further emphasise the importance of this topic, Overett and Donald (1998) highlight literacy's ability to equip people with cultural power and identity, which may – particularly for those with marginalised identities – give individuals the necessary skill set to transform their life. There is a large body of research confirming how early reading ability holds predictive validity for later child development and academic outcome (Du Plessis et al. 2003; Howie et al. 2008; The National Institute for Literacy 2008; Spaull 2011). It has been shown that interventions targeted at lower primary school level can be very effective in serving as relatively swift corrective measures for those who struggle to read (He et al., 2009; The National Institute for Literacy 2008).

Undeniably, the current South African educational system has failed to deliver. Sub-standard academic performance and other education metrics, such as high dropout rates and grade retention, indicate a rather dire situation.[2] This is in large part due to the structural and systemic issues that are deeply entrenched in the post-apartheid South African schooling environment. Apartheid left its footprint on the South African educational system, with unequal access to education and varying levels of educational quality – when it comes to both private and social returns on schooling, disparities along racial lines still exist. Even though the

---

[1] By being functionally illiterate, a learner is unable to read a simple, short text and infer any meaning from the written text (Spaull 2011).

[2] According to the *Education Statistics in 2014* report, with specific reference to the Western Cape, only 6.2% of Grade 9 learners achieve a score of 50% or higher in mathematics, the average score being 13% (Department of Basic Education [DBE] 2016). Similarly, the average score for first additional language (which in most cases is English) is 38%, with only 24.5% scoring 50% or higher – slightly better than mathematics performance, but still dramatically sub-par.

lives of African and Coloured people have generally improved, returns to education remain much higher for white people, who earn 40% more than Africans and 20% more than Coloured individuals per added year of schooling (Salisbury 2016).

The Shine Literacy programme offers an important, high-quality input to resource-poor schools, with the goal of addressing some of these educational shortfalls and thus redressing injustices of the past early on. Shine Literacy is a South African non-profit organisation (NPO) that provides reading assistance to learners who are identified to be at literacy risk by the end of Grade 1. Assistance takes place during Grade 2, in the form of bi-weekly one-on-one sessions between volunteers and at-risk learners, teaching phonetics, word recognition and decoding, and the shared reading of whole paragraphs. If the learner does not improve sufficiently after the completion of Grade 2, they will be taken on for another year during Grade 3. Shine's programme is embedded in the mutual understanding between government, business and NPOs that South Africa's literacy problem requires strong collaboration between a variety of actors, calling for a holistic approach. Shine Literacy aims to ensure that learners are equipped with their developmental age/grade adequate literacy level. Their ethos makes it clear that early interventions imply large gains in terms of academic self-esteem in later years, keeping the goal of promoting a culture of learning in mind. This mission follows the notion emphasized in the work on the returns to education by Heckman (2006; 2011) and Heckman and Masterov (2004), namely, that the earlier the intervention, the more persistent and socially just the returns will be. For instance, Heckman and Masterov (2004) estimated that the rate of return of an early intervention program designed for low-ability children is 16%.

This study evaluates the impact of Shine Literacy by use of a difference-in-differences approach.[3] The analysis can be placed in the international discussion on which inputs to education are required to overcome the persistent lack of quality. With a potential lookout for externality effects, this study did not only aim to measure the programme impact on learners' literacy, but also to provide some evidence on why Shine Literacy should be scaled-up across South Africa. Large standard deviation literacy improvements of up to 1.9 were found, with the most conservative estimate starting from 0.6. This amounts to educationally meaningful results. Further, it appeared that those who were most deficient with regards to their literacy and thereby later schooling were assisted the most, and experienced the greatest literacy improvements. No heterogeneous effects on gender were found when assigned to the programme, boy learners do just as well as girl learners – this is an important insight into closing the gender gap that exists in literacy.

By way of introduction, the following section discusses the available data and methods used to evaluate the Shine programme, as well as the results of the analysis and their contextual implications. By way of conclusion, the discussion section aims to provide some linkages to the international evidence and deliver not only some remarks on the limitations of the analysis, but also some suggestions as to future work.

2 THE SHINING STARS OF SHINE LITERACY

Shine Literacy's first pilot centre opened in 2000 at the Observatory Junior School in Cape Town, with residents of the suburb acting as 'reading partners' for Grade 2 or Grade 3 learners identified by their teachers as requiring reading assistance (The DG Murray Trust 2012). After positive results,[4] confirmed by the Western Cape Education Department (WCED), Shine Literacy officially registered as an NPO, opened more centres at various schools across Cape Town during the late 2000s, and formalised its programme (The DG Murray Trust 2012). In essence, the programme's strategy is to provide primary schools with trained volunteers that assist learners during normal school hours. This assistance takes place in the form of one-on-

3 The dissertation from which this paper is drawn also includes the approach of propensity score matching.

4 The WCED Grade 3 Systemics had shown a continuous increase in literacy scores at Observatory Junior School, starting from 50% in 2002, up to 71.1% by 2006 (The DG Murray Trust 2012).

one sessions, for two hours between volunteers and Grade 2 learners twice every week (The DG Murray Trust 2012).

Regarding its content, the programme offers access to storybooks and other reading material, both in English and the learners' home language, if that is different. During the bi-weekly sessions, a Shine-trained volunteer works together with a learner to improve their phonetic understanding, word recognition and decoding, as well as the paired reading of whole paragraphs. Moreover, Shine invites learners' caregivers to two workshops during Grade 1, before the commencement of the programme, which provides them with the necessary resources and information to scaffold their children's reading progress at home (Shine Literacy 2016). Shine's goal is to offer an important high-quality input into resource-poor schools and address educational shortfalls. Moreover, the programme aims to ensure that Grade 2 and 3 learners are equipped with their developmental age/grade adequate literacy (The DG Murray Trust 2012; Shine Literacy 2016). This is emphasised by the recognition that these grades are crucial in guiding literacy acquisition and that, by Grade 4, learning to read will become increasingly difficult (The DG Murray Trust 2012).

At Shine Literacy, the literacy ability of all children in the classroom is assessed at the end of Grade 1 with the same test instrument across all participating schools (The DG Murray Trust 2012). This test diagnostic[5] contains a battery of various literacy assessment items and assesses learners' ability to identify alphabet sounds, write down three-letter words after being given picture hints, and write a sentence that is dictated to them. Shine has identified these skills as essential in learning how to read for meaning (Shine Literacy 2016). After taking the test, learners are sorted into a literacy proficiency ranking, categorised as either 'At risk', 'Poor', 'Satisfactory' or 'Good', with regard to their abilities. The following test scores are associated with the rankings: 0–30% (At risk); 31–59% (Poor); 60–70% (Satisfactory); and 71–100% (Good) (Volbrecht personal communication 2018, January 30). In practice, programme assignment does not clearly follow the underlying assumption behind this ranking. Because of varying degrees of availability among volunteers at the centres, some learners that are outside the 'At risk' or 'Poor' category are often assigned to the programme; that is, when there are more volunteers than otherwise needed, rather than sending them away, Shine assigns a learner to them. In the same vein, often too few volunteers are available, and some 'At risk' or 'Poor' learners do not get assigned a volunteer (Dornbrack personal communication 2017, January 20). Mostly, the concern is that there are too few, rather than too many, volunteers on the school roster. Nevertheless, the Programme Manager for Shine Centres reports that once the roster is established for the year, volunteer numbers seem to stay constant over the school terms, with only minor fluctuations. The number of volunteers across all Shine schools is currently around 770–800 (Volbrecht personal communication 2018, February 6).

By the middle of Grade 2, a second test diagnostic is taken by all learners, both non-participants and participants. This is to measure whether the latter have made sufficient progress to leave the programme, or whether they should stay in the programme for another year, during Grade 3. The programme does not currently extend beyond Grade 3 and all Shine participants go back to the general learner population (The DG Murray Trust 2012). Generally, the progress of all Shine and non-Shine learners during the whole of Grade 2 is monitored, and feedback is given to the teacher, parents, heads of department and schools (The DG Murray Trust 2012; Shine Literacy 2016).

5 The diagnostic was designed by three South African educationists in continuous collaboration with centre staff and external Shine partners such as the South African NPO Wordworks. The diagnostic itself has not been psychometrically validated, due to funding constraints. However, Shine has emphasised that the diagnostic itself has to be seen as a formative assessment, as opposed to a summative assessment (Volbrecht personal communication 2018, January 30). More on the different functions and uses of these two forms of assessments can be found in the *Guide to Assessment in Early Childhood: Infancy to Age Eight* by Slentz et al. (2008).

A previous evaluation of Shine was conducted by Schkolne (2014), who used data from eight Shine participating locations in the Western Cape for the years 2011–2013. Schkolne used a sharp cut-off discontinuity regression to show that the test score targets[6] set by Shine were met or in many cases exceeded. This includes the fact that, after half a year, most Shine participants' results increased to the 77.77% grade level average and was shown to be a statistically significant result of attending Shine even after 12 months at 10% significance. However, Schkolne found that this effect diminished after 18 months and that learner improvement was not a direct result of having attended Shine.[7]

2.1 Data and an introduction to the sample

To evaluate the impact of Shine on literacy, this analysis made use of three datasets: Shine baseline and endline test score data; Shine attendance data; and Systemic test scores and demographic data from participating schools obtained from the WCED for 2013 to 2016. By combining the Shine and WCED data using the unique Central Education Management Information System identifier, it was possible to fill in missing information in the Shine datasets from WCED data. This included information on sex, birthdate and home language of the learners. The instruction (i.e. the way it is assessed) and content (i.e. what is assessed) of Shine's test instrument had remained constant between 2013 and 2016, which is the time period underlying this analysis[8]. In that way, learner progress can be compared across and within cohorts, given the constant nature of the test content. Pooling the available Shine test score data and dropping observations with missing test scores resulted in a dataset of n=1543 unique observations. This dataset contains information on Shine participants and non-participants across ten schools over four years. Programme participation was recorded by Shine with a categorical variable indicated as 'Never attended',[9] 'Participating', 'Graduated', 'Waiting' or 'Special needs'. In the data-cleaning process, learners listed as 'Graduated' were re-coded as participants. Similarly, learners indicated as 'Waiting' (n=12) were also re-coded as participants, given the intention to treat those individuals. Lastly, learners that were identified as requiring special education (n=5) were dropped from the sample. This resulted in a sample of 836 non-participants and 707 participants.

In the Western Cape, ten primary schools had hosted[10] an independent Shine Centre by 2018, all in the metro area of the City of Cape Town. Table 1 gives a breakdown of the participating schools, the total number of Grade 2 pupils, the fraction enrolled at Shine, data availability by year from each school, as well as Shine attendance rates.

6 The specific target mentioned in Schkolne (2014) refers to Shine's first medium-term goal of improving each Shine participant's score by 20% (Grigg et al. 2016).

7 Interpreting Schkolne's (2014) results needs to be done with caution, since Shine does not strictly adhere to the cut-off points for selection. Therefore, the cut-off used in her work may result in – at best - a fuzzy discontinuity.

8 Note that Shine has adopted a new testing diagnostic, which has been piloted since 2017.

9 To avoid confusion: this is how Shine labelled those learners that were not assigned to the programme, i.e. non-participants.

10 On the process of school selection, Shine identified their hosting schools based on the following school/learner characteristics: little to no formal pre-school education for learners; little available remedial support; teacher-learner ratios; little access to books or libraries; and mostly second language learners from low-income households (The DG Murray Trust 2012; Shine Literacy 2016).

Table 1: Participating schools

School	Total Grade 2 pupils (n)	Shine enrolment (%)	Years data availability				Attendance rates (%)
			2013	2014	2015	2016	
Claremont Primary	123	52.8	x	x	x	x	77.9
Good Hope Seminary Junior	177	41.8	x	x	x	x	55.9
Observatory Junior	68	77.9	x	x			79.3
Prestwich Street Primary	264	42.4	x	x	x	x	71.2
Rosmead Central Primary	159	38.9			x	x	71.5
St. Agnes's Primary	174	43.6		x	x	x	54.3
St. Paul's Primary (Wynberg)	252	40.0	x	x	x	x	55.0
Walmer Estate Primary	96	48.9		x	x	x	59.4
Zonnebloem Boys Primary	128	61.7	x	x	x	x	53.4
Zonnebloem Girls School	102	35.2	x	x	x		64.2
TOTAL	1 543	45.8					63.5

The largest school in the sample was Prestwich Street Primary with 264 Grade 2 learners, followed by St. Paul's Primary with 252. Much smaller was Observatory Junior with 68 pupils, and Walmer Estate Primary with 96. Across all schools, Shine enrolment was 45.6% on average. The highest fraction of Shine participants was found in Observatory Junior (site of the pilot Shrine centre), where 77.9% of Grade 2 pupils were enrolled in the programme. This is followed by 61.7% of learners enrolled at Zonnebloem Boys Primary and 52.8% at Claremont Primary. In contrast, at Zonnebloem Girls only 35.2% of Grade 2 learners were enrolled. All ten schools were identified as falling into the NQ5 Quintile. All teach exclusively in English.

There are attendance data available for all 707 Shine participants. Learner attendance at Shine sessions are generally recorded on a monthly basis. This was aggregated for each learner over the whole school year. The result is a list of information on how many sessions there were available for each learner and, out of that, how many sessions the learner actually attended. The resulting intensity index measures the fraction of available sessions attended by the learner. On average, Shine learners attended 63.5% of their sessions. This varies by school: the highest average attendance was recorded at Observatory Junior, where Shine learners attended almost 80% of their sessions. In contrast, participants at Zonnebloem Boys Primary attended 53.4% of their sessions.

To get a deeper understanding of the available data, and to assess the internal validity of the analysis, Table 2 and Table 3 take a more detailed look at the sample characteristics. Table 2 displays the differences in demographics between participants and non-participants to examine the balance in the two sample populations. The overall sample was split almost equally between boys and girls, but boys were significantly more likely to be in the participant population. Across languages, both samples seem to be similar: there are no statistically significant differences, with the exception of IsiXhosa speakers, who were more likely to be participating in the programme. Regarding learner age, no differences exist. On average, a learner was 8.5 years old. The majority (29.8%) of the sample spoke IsiXhosa, which was followed by English (12.7%).

Table 2: Sample demographics (%)

		%	Participant	Non-participant
All (n, %)		1543 (100)	707 (45.8)	836 (54.2)
Sex	*Female*	50.8	38.7	61.3***
Language[a]	*IsiXhosa*	29.8	51.1	48.9***
	English	12.7	47.7	52.3
	French	0.8	25	75
	Afrikaans	0.5	57.1	42.9
	Other[b]	3.2	56	44
Age (mean, SD)		1543	8.51 (0.5358)	8.48 (0.4309)

Notes:

a. Given that language data were imputed from the Systemic dataset and not all learners were able to be matched with Systemic information, there were a total of 821 missing data on home languages.

b. This category included IsiZulu, IsiNdebele, Pedi, Sesotho, Tswana, Venda, and Xitsonga.

c. *** p<0.01, ** p<0.05, * p<0.1

To gain an idea of how test scores varied between participants and non-participants, Table 3 shows pooled test score averages for the Shine and Systemic tests, by participation. At baseline, Shine pupils scored an average of 44.4% on the Shine test diagnostic, compared to 70.5% scored by non-participants. This difference is statistically significant. At endline, Shine pupils' test score average increased to 66.4%, but their scores remained statistically lower than non-participants who averaged a score of 74.8 %. When looking specifically at the literacy ranking assigned at baseline, the vast majority of Shine participants scored either in the 'At risk' (27%) or 'Poor' (49.4%) categories. 19.2% of those that were assigned to the treatment scored 'Satisfactory' and 4.4% 'Good'. In comparison, the vast majority (65.6%) of non-participants were ranked as 'Good'. This was followed by 'Satisfactory' learners who made up just over 30% of the non-participant population. Only 2.5% and 1.8% of the non-participants scored in the 'At risk' and 'Poor' categories, respectively.

Table 3: Shine test scores

			n	Average all (%)	Participants (%)	Non-participants (%)
Shine test score	*Baseline (Grade 1)*		1 543	58.5	44.4	70.5***
	Literacy categories, %	*At risk*	212		27	2.5
		Poor	364		49.4	1.8
		Satisfactory	388		19.2	30.1
		Good	579		4.4	65.6
	Endline (Grade 2)		1 543	70.9	66.4	74.8***
	Literacy categories, %	*At risk*	46		3.5	2.2
		Poor	130		17.3	1
		Satisfactory	269		29.4	7.3
		Good	1 094		49.8	89.6

*** p<0.01, ** p<0.05, * p<0.1

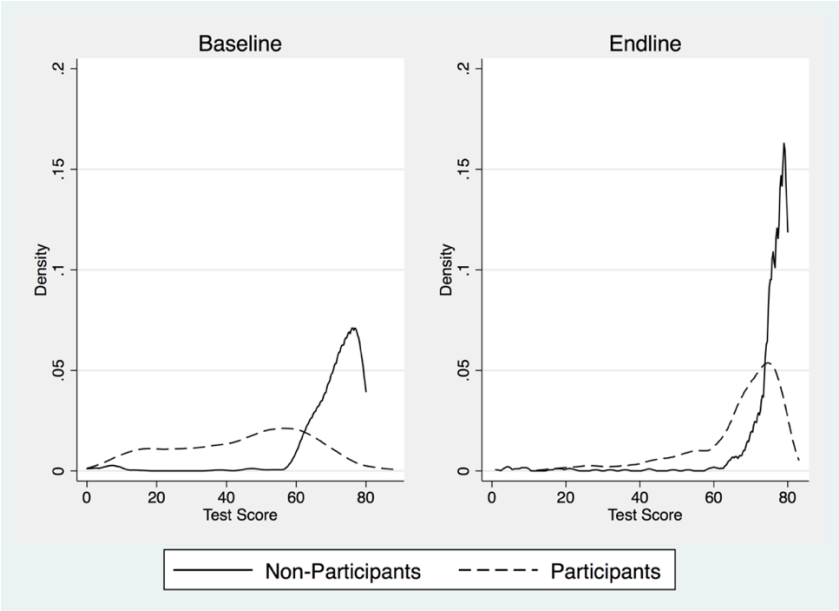
At endline, the majority (49.8%) of Shine participants ranked in the 'Good' literacy category. This is followed by 29.4% that scored 'Satisfactory'. Only 3.5% of Shine learners scored in the 'At risk' category and 17.3%

scored in 'Poor'. In the case of non-participants, 89.6% of learners were ranked as 'Good', whereas 7.3% fell in the 'Satisfactory' range. Lastly, 1% of learners who were not assigned to the programme scored in 'Poor', and 2.2% were considered 'At risk'.

2.2 Shine's diagnostic: How do participants compare to non-participants?

As highlighted by Ho and Yu (2015), normality in test score distributions rarely holds in practice. A quick Skewness and Kurtosis hypothesis test for normality on Shine scores showed that neither the baseline scores nor the endline scores are distributed normally. Instead, the distributions are heavily skewed and have much thinner tails[11] than a normal distribution. This is illustrated in Figure 1, which maps out the distribution of Shine test scores at baseline and endline by treatment. At baseline, non-participants' scores are much higher and concentrated around 70%, compared to a more flattened and spread out score distribution among participants. By endline, however, both samples' tails become tighter, meaning that not only do the score means increase, the distribution becomes less spread out. This decreases the variance in learner performance and essentially makes instruction for the teacher much easier. To really observe the magnitude of these shifts in the two distributions by participation, Figure 2 shows specifically how participants' scores change from baseline to endline, compared to how non-participants scores shift. Participants' score distribution moved from its spread-out nature during baseline to a much tighter distribution at endline. A much higher fraction of Shine learners was found above the 'Satisfactory' cut-off line. Recall that just under 30% scored in 'Satisfactory' and around 50% in 'Good'. As such, mean test scores were still lower non-participant learners' scores. However, the magnitude of this mean increase among Shine learners is striking and is already pointing towards large difference-in-differences

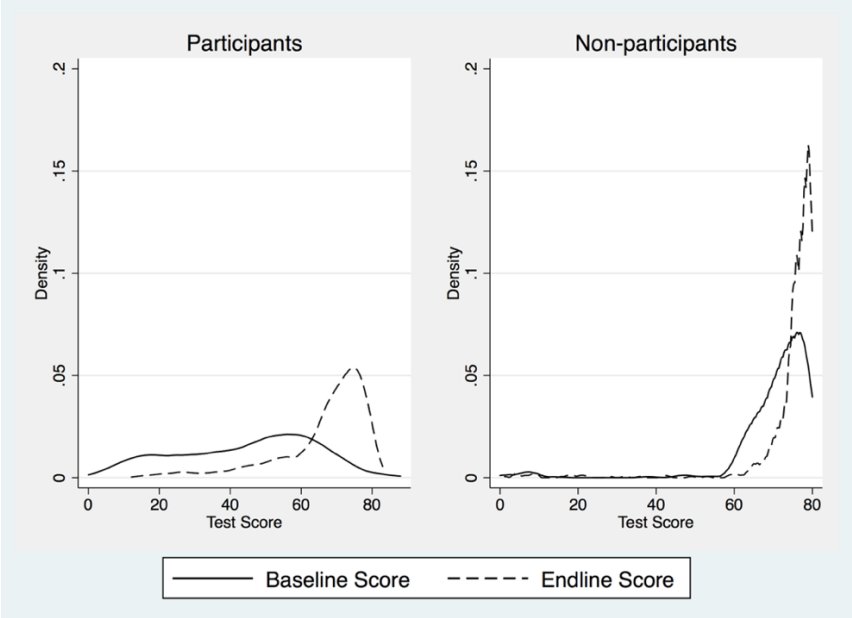
Figure 1: Shine test score distribution over time



Note: Baseline non-participants' kurtosis and skewness: 22.02 and –3.04, participants' kurtosis and skewness: 2.12 and -0.3; endline non-participants' kurtosis and skewness: 32.76 and -5.21; participants' kurtosis and skewness: 5.82 and -1.72.

11 Hypothesis p-values: Skewness Baseline: p=0.000, Kurtosis Baseline: p=0.01312; Skewness Endline p=0.000, Kurtosis Endline: p=0.000.

Figure 2: Shine test score distribution by participation



2.3 Difference-in-differences: 'At risk' learners and isiXhosa speakers take it home

The Shine test score data lend itself to the use of difference-in-differences techniques, which has become a widespread method in quasi-experimental studies where assignment to treatment is not randomised (Gertler et al. 2012; Khandker et al. 2010; Lance et al. 2014). The most straightforward set-up is to observe the same test diagnostic for both non-participants and participants before and after the treatment. Non-participants were not assigned to the treatment, but participants were. To remove any bias in the endline comparison between the participants and non-participants that are either due to permanent differences between the groups or the result of time-variant trends,[12] the difference-in-differences method subtracts the average difference in the non-participant group from the average difference in the participant group. Given the Shine data, the performance of all Shine participants is then compared to all non-participating learners, where both populations were spread across all four literacy ranking categories. This made difference-in-differences estimation an attractive method to apply.

First, consider the difference-in-differences equation if the data were treated as repeated cross sections. Shine treatment status was indicated by S = 0, 1 where 0 indicated no treatment, i.e. non-participants or the comparison group, and S = 1 indicated a Shine learner. Further, all learners were observed over two time periods, t =0, 1 where 0 was the time period at baseline, and 1 was the period at endline. Every learner was indexed by the letter i =1, …N and learners have two observations. Then, $\overline{\text{testscore}}_0^{\,S}$ and $\overline{\text{testscore}}_1^{\,S}$ are the average outcomes in test scores for Shine learners at baseline and endline, respectively. For non-participants the average outcomes were $\overline{\text{testscore}}_0^{\,NS}$ and $\overline{\text{testscore}}_1^{\,NS}$. The difference-in-differences are then as shown in Equation 1:

$$\zeta=(\overline{\text{testscore}}_1^{\,S}-\overline{\text{testscore}}_0^{\,S})-(\overline{\text{testscore}}_1^{\,NS}-\overline{\text{testscore}}_0^{\,NS}) \quad (1)$$

12 Time-variant trends would be the tendencies of test scores to improve generally as learners proceed through the grade, in line with age developmental progress.

Manually taking these differences across the available independent variables, resulted in a list of difference-in-differences. This is shown in Table 4, which displays difference-in-differences across sex, home language, schools, as well as the baseline literacy categories.

Table 4: Test score difference-in-differences (%)

		Participants			Non-participants			ζ
		Baseline	**Endline**	**Difference (x)**	**Baseline**	**Endline**	**Difference (y)**	**x - y**
All		44.4	66.4	22	70.5	74.8	4.3	17.7***
Sex	Female	47.7	68.7	21	71.7	75.8	4	16.9***
	Male	41.4	64.2	22.8	68.5	72.8	4.3	18.4***
Language	isiXhosa	41.1	66.2	25.1	70.9	74.7	3.8	21.5***
	English	52.3	67.9	15.6	72.2	77.1	4.9	10.6***
	French	50.3	70.4	20.1	74.5	78.3	3.8	16.3***
	Afrikaans	41.5	62.2	29.7	74.7	77.7	3	17.7
	Other	47.9	67.9	20	71.9	76	4.1	15.9***
School	Claremont Primary	58.6	75.7	17.1	75.9	78.1	2.2	14.9***
	Good Hope Seminary Junior	45.4	68.6	23.3	73.4	75.9	3.5	19.8***
	Observatory Junior	35.1	68.7	33.6	52.3	57.6	5.3	28.3***
	Prestwich Street Primary	45.7	67.7	22.5	69.9	73.7	3.8	18.7***
	Rosmead Central Primary	57.1	72.6	15.5	75.8	73.7	2.3	13.2***
	St. Agnes's Primary.	51.2	62.9	11.7	70.8	76.6	5.8	5.9***
	St. Paul's Primary	32	54.2	22.2	64.1	69.6	5.5	16.7***
	Walmer Estate Primary	36.9	63.5	27.6	68.7	73.4	4.7	22.9***
	Zonnebloem Boys Primary	36.9	65.3	28.3	68.1	74.1	6	22.3***
	Zonnebloem Girls Prac. Sch.	49.7	70	20.4	73.6	76.8	3.2	17.1***
Baseline literacy	At risk	18.6	53.2	34.6	5.8	15.7	9.9	24.7***
	Poor	47.4	69.7	22.3	51.1	69.9	18.8	3.5
	Satisfactory	65.5	73.1	8.6	66.1	73.8	7.7	0.9
	Good	76.5	75.1	-1.9	75.6	77.5	1.9	-3.3***
	*** p<0.01, ** p<0.05, * p<0.1							

The difference-in-differences for all categories were relatively large throughout, and with the exception of the Afrikaans, 'Poor', and 'Satisfactory' categories, all are statistically significant. The difference-in-differences for Shine participants on test scores indicated an increase by about 17.7% due to attending Shine. The largest effect across languages among Shine participants was found for isiXhosa speakers with a difference-in-differences of 21.5%. When examining the schools, the highest test score difference-in-differences was recorded in Observatory Junior with 28.3%; followed by Walmer Estate Primary with 22.9%. The lowest difference-in-differences was found for St. Agnes's Primary with 5.9%. Lastly, across literacy rankings, 'At risk' Shine learners saw a difference-in-differences in their test scores of 24.7%. In comparison, 'Good' participants' scores actually saw a negative difference-in-differences of 3.3%.

Making use of the two-time period structure of the Shine test score dataset, it was possible to estimate the following model and arrive at a consistent estimator of the average treatment effect of Shine, as in Equation 2:

$$\Delta \text{test score}_{it} = \alpha + \beta S_{it} + \gamma\ \text{testscore}_0 + \delta(S_{it} * \text{testscore}_0) + \phi x_{it} + \varepsilon_{it} \quad (2)$$

where
β = measures the Shine-specific average treatment effect;
γ = captures the impact of baseline test scores;
δ = measures the effect of the interaction between Shine treatment and baseline test scores;
ϕ = captures the effect of a vector of independent characteristics, including school dummies; and
ε = an idiosyncratic, unobserved error term.

To see how robust the coefficient on the treatment indicator was, this model included covariates of gender, learner's age and age squared (to account for a possible quadratic effect), learner's home language, schools and learner's attendance, including a squared attendance term. Robust standard errors were used and clustered at the school level. The results of this estimation are displayed in Table 5. In addition to the estimation of the average effect on the whole learner sample, various interactions were also included, in order to observe a potential avenue for differential effects. The effect of those interactions is added as additional columns in the table.

Table 5: Change in test scores as a function of Shine

Dependent variable: Change in Shine test score	(1)	(2)	(3)	(4)	(5)	(6)
Shine	26.94*** (2.381)	26.85*** (2.482)	24.94*** (2.552)	4.592* (2.070)	3.942 (2.397)	-3.717 (3.676)
Shine baseline score	-0.240*** (0.0574)	-0.241*** (0.0576)	-0.260*** (0.0597)	-0.490*** (0.0904)	-0.483*** (0.0943)	
Female				1.688*** (0.469)		
Shine X female				0.120 (1.320)		
Learner's age		13.31 (15.51)	10.29 (16.29)	13.23 (17.70)	13.83 (17.41)	6.788 (17.25)
Learner's age squared		-0.683 (0.888)	-0.534 (0.926)	-0.702 (1.000)	-0.742 (0.988)	-0.327 (0.985)
Shine X baseline score	-0.347*** (0.0405)	-0.345*** (0.0392)	-0.326*** (0.0371)			
isiXhosa					-0.841 (0.701)	
Shine X isiXhosa					2.382 (1.428)	
At risk category						9.980*** (1.537)
Poor category						16.65*** (1.701)
Satisfactory category						6.510*** (0.759)
Shine X Poor						7.055* (3.752)
Shine X At risk						27.01*** (4.337)
Shine X Satisfactory						2.844 (2.605)
Good Hope Seminary Junior			-0.572 (0.334)	0.0104 (0.663)	-0.0652 (0.705)	-1.463*** (0.311)
Observatory Junior			3.634** (1.311)	4.082** (1.752)	3.694** (1.464)	3.134*** (0.769)

Dependent variable: Change in Shine test score	(1)	(2)	(3)	(4)	(5)	(6)
Prestwich Street Primary			-1.335**	-1.151	-1.315	-2.640***
			(0.514)	(0.918)	(0.803)	(0.474)
Rosmead Central Primary			-1.528***	-0.966**	-1.071***	-1.495***
			(0.219)	(0.311)	(0.257)	(0.378)
St. Agnes's Primary			-3.278***	-3.269***	-3.276***	-5.288***
			(0.302)	(0.616)	(0.808)	(0.378)
St. Paul's Primary			-4.067***	-4.048**	-4.200**	-4.653***
			(0.908)	(1.643)	(1.770)	(0.421)
Walmer Estate Primary			0.0599	0.572	0.457	-1.391*
			(0.851)	(1.414)	(1.321)	(0.631)
Zonnebloem Boys Primary			1.002	2.412	1.352	-0.327
			(0.956)	(1.762)	(1.316)	(0.477)
Zonnebloem Girls Prac. Sch.			-0.952**	-1.205**	-0.456	-2.068***
			(0.315)	(0.438)	(0.460)	(0.509)
Constant	21.15***	-42.41	-24.63	-22.43	-23.89	-29.74
	(4.132)	(66.12)	(69.72)	(75.52)	(74.00)	(74.88)
Observations	1 543	1 543	1 543	1 543	1 543	1 543
R-squared	0.699	0.702	0.719	0.695	0.693	0.691

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Estimating Equation 2 resulted in an average treatment effect ranging between 24.9% and 26.8%, depending on the specification. These coefficients are statistically significant at 1% and robust throughout. In practical terms, this implies that the effect of being assigned to the programme increased test scores between 1.7 and 1.9 standard deviations. Compared to other education programmes, this is already an enormous test score increase. An impact of this magnitude falls in the range of large Cohen's *d* effect sizes, which is implied by anything greater than 0.8 standard deviations (Cohen 1992). Educationally speaking, programmes that increase scores by over 0.2 standard deviations are already viewed as meaningful (Cohen 1992; Wasik 1998).

Although girls across the learner population generally did around 1.7% better at endline, the coefficient on the interaction of Shine treatment and sex is small and statistically not different from zero. This implies that Shine impacts all learners the same, regardless of their sex. In terms of age, though there was the expected quadratic effect (impact of age decreases as age increases), in none of the specifications was the learner's age statistically significant. Unsurprisingly, the lower the learner's baseline score, the lower they scored at endline. This holds for both the general population and the Shine learners. Including an isiXhosa dummy carried no effect – presumably because of fixed school effects.[13] From column 6 in Table 5, it can be seen that large positive gains are experienced by 'At risk' learners with a test score improvement of 27%, which is statistically highly significant at 1%.

To gain more insight on the potentially heterogeneous effects that Shine has on its learners, and to see who benefitted the most from participating, the following regression procedure allowed for the estimation of the differential impact on various sub-samples. The results of this estimation are displayed in Table 7, where in each column the same specification as per equation (2) above was run but restricting the sample on the relevant population.

13 This is probably indicative of different demographics represented at each school; in this case, some schools are more heavily skewed toward isiXhosa speakers. For more insights on how school differ by demographics, see Appendix A.

Table 6: Differential impact of Shine by baseline literacy profile

Dependent variable: Change in Shine test score	(1) Sample: Girl children only	(2) Sample: isiXhosa children only	(3) At risk learners	(4) Poor learners	(5) At risk & Poor
Shine	20.09*	24.14***	20.79**	12.03	34.31***
	(9.231)	(6.789)	(6.501)	(7.608)	(4.721)
Shine baseline score	-0.381**	-0.296***	-0.592	-0.486***	0.124**
	(0.142)	(0.0889)	(0.648)	(0.137)	(0.0464)
Shine * baseline score	-0.261*	-0.296**	0.349	-0.221	-0.619***
	(0.127)	(0.0967)	(0.761)	(0.143)	(0.112)
Good Hope Seminary Junior	-0.300	-1.156*	-6.191***	-3.116**	
	(0.495)	(0.549)	(0.114)	(1.218)	
Observatory Junior	1.659	4.045***	0.948	-0.522	
	(0.923)	(0.952)	(0.747)	(1.155)	
Prestwich Street Primary.	-1.427**	-1.537***	-6.303***	-4.687***	
	(0.609)	(0.303)	(0.574)	(1.194)	
Rosmead Central Primary	-1.099**	-2.482***	2.531***	-3.984**	
	(0.439)	(0.129)	(0.555)	(1.682)	
St. Agnes's Primary	-2.809***	-4.569***	-14.71***	-9.818***	
	(0.587)	(0.219)	(0.102)	(1.799)	
St. Paul's Primary	-2.015	-1.011	-14.94***	-9.804***	
	(1.237)	(0.873)	(1.097)	(1.525)	
Walmer Estate Primary	0.756	-1.977**	-3.893***	-3.836***	
	(0.789)	(0.666)	(0.314)	(1.210)	
Zonnebloem Boys Primary		-2.089**	-0.821	-4.466***	
		(0.706)	(0.539)	(1.118)	
Zonnebloem Girls Prac. Sch.	-0.893	-1.146***	-0.510**	-2.657**	
	(0.493)	(0.244)	(0.204)	(1.240)	
Constant	32.64**	26.45***	25.74***	48.77***	10.99***
	(10.60)	(6.640)	(3.384)	(7.334)	(1.389)
School FE					YES
Observations	781	460	212	364	576
R-squared	0.748	0.788	0.340	0.508	0.362

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

In the first column of Table 6, it was shown how girl learners perform by participation: that is, to see how Shine girls fare against non-Shine girls. While there was a 20.1% effect increase, this is only significant at the 10% level. In contrast, restricting the learner population to isiXhosa speakers only, it became clear that they benefitted greatly from the programme: test scores increased by 24.1% when assigned to the programme – an effect that is statistically significant at 1%. Furthermore, across baseline literacy profiles, the 'At risk' learners were the strongest beneficiaries. Test scores among those who most struggled with literacy increased by 20.8% as a result of participating in the programme. This effect increase is significant at the 5% level. It is, however, important to note that the sample for this estimation decreased down to 212 learners. Of these, the majority were Shine participants, which led to the problem of small sample comparison. To mitigate this problem, the specification was run on a pooled sample of 'At risk' and 'Poor' learners and its results reported in column 5. The treatment effect here was strong and statistically significant, even with school fixed effects. In terms of schools generally, Good Hope, Prestwich Street, St. Agnes's, St. Paul's, Walmer Estate, Zonnebloem Boys and Zonnebloem Girls all did statistically worse than Claremont Primary, which is the omitted school across the included school dummies.

2.4 Transitions to better literacy

To get an understanding of how learners improved between baseline and endline, Table 7 shows how the general learner population transitioned along the literacy rankings from baseline to endline. The way this transition varied by Shine participation is shown in Table 8.

Table 7: Literacy categories all (%)

All learners		Grade 2			
		At risk	Poor	Satisfactory	Good
Grade 1	At risk (n=212)	19.3	42.9	22.1	16.5
	Poor (n=364)	0.3	9.1	37.1	53.6
	Satisfactory (n=388)	0.3	1	19.9	78.9
	Good (n=579)	-	0.4	2.1	97.6

In general, fewer than 20% of baseline 'At risk' literacy learners stayed in this category by endline, 42.9% moved up to 'Poor', 22.1% to 'Satisfactory' and 16.5% to 'Good'. 'Poor' learners at baseline were generally not seen again in this category: under 10% remained there by endline. In contrast, 37.1% moved to 'Satisfactory' and more than half scored in the 'Good' range. Out of those learners that were categorised as 'Satisfactory' at baseline, just under 20% stayed in 'Satisfactory' and the majority (78.9%) scored 'Good'. Lastly, the overwhelming majority (97.6%) of 'Good' literacy learners stayed in 'Good'. Only 2.1% slid down to 'Satisfactory', and a small fraction of 0.4% down to 'Poor'.

Table 8: Literacy categories by participation, %

Participants		Grade 2				Non-participants		Grade 2			
		At risk	Poor	Satisfactory	Good			At risk	Poor	Satisfactory	Good
Grade 1	At risk (n=191)	12.6	45.6	23.7	18.3	Grade 1	At risk (n=21)	81	19	-	-
	Poor (n=349)	0.3	8.9	37.5	53.3		Poor (n=15)	-	13.3	16.7	60
	Satisfactory (n=136)	-	2.2	21.3	76.5		Satisfactory (n=252)	0.4	0.4	19.1	80.2
	Good (n=31)	-	3.2	9.7	87.1		Good (n=548)	-	0.2	1.6	98.2

When looking at the transition by participation displayed in Table 8, only 12.6% of those learners who were categorised as 'At risk' and assigned to the programme remained there. In contrast, 45.6% moved to 'Poor', 23.7% to 'Satisfactory' and 18.3% moved up to 'Good'. Similarly, around 9% of learners who scored 'Poor' at baseline stayed there. In fact, 37.5% moved up one category to 'Satisfactory' and the majority (53.3%) increased by two categories to 'Good'. 'Satisfactory' learners who were assigned to Shine at baseline stayed in that category 21.3% of time, while the overwhelming majority of 76.5% moved up to 'Good'.

On the other hand, non-participant learners that were considered 'At risk' (n=21) remained in that category 81% of the time. The remaining 'At risk' fraction moved up to 'Poor', however, none moved to 'Satisfactory' or 'Good'. This essentially implies that those 'At risk' learners that were not assigned to the programme, stayed stuck. They did not transition to better literacy by endline. However, it needs to be highlighted that the sample size of those 'At risk' learners is very small to begin with. Another small, yet striking insight into how learners transition is found in the 'Good' category: a higher proportion (98.2%) of the non-participants who were considered 'Good' remained in 'Good' when compared to participants (87.1%); however, this only pertained to n=31 participant cases. In the instances of 'Good' participants, roughly 10% actually moved

down to 'Satisfactory'. This compares to only 1.6% of 'Good' non-participants (n=548) who moved down to 'Satisfactory'. Admittedly, the drastically smaller sample of 'Good' learners that participated against the much bigger sample of 'Good' non-participants, does not give rise to any serious concern about this movement.[14]

3 SHINE LITERACY AS THE SOUTH AFRICAN COUNTERPART TO SHISHUVACHAN CLASSES

Volunteering in education programmes has increased momentum over the last two decades (see for instance, Wasik (1998); Elliott et al. (2000) and Porter & Johnson (2004)), and it does not only seem to be causing a 'warm glow' among volunteers, but also holds educational promise for the learners. Shine Literacy's method, one that is largely based on interactions between volunteers and learners, serves as a great example, which seems to be working well. Across educational interventions,[15] the impact Shine Literacy has on its learners' performance is much greater in magnitude. Complex education programmes such as conditional cash grants, as evaluated by Baird et al. (2011), hiring additional contract teachers as analysed by Duflo et al. (2015), or providing school vouchers as examined in Angrist et al. (2006), resulted in small standard deviation improvements of between 0.12 and 0.29. Looking more specifically at literacy programmes, the large standard deviation upwards of 1.7 standard deviations observed for the Shine Literacy case is unparalleled. Teachers using print referencing techniques as evaluated by Justice et al. (2009), or applying ICT methods in the classroom as examined in Brooks et al. (2006), caused medium-sized effect changes of around 0.5 standard deviations in participant learners' literacy. In contrast, Shine Literacy can almost compete with the famous Shishuvachan classes evaluated by He et al. (2009). Children assigned to the Shishuvachan sessions improved their scores by up to 0.7 standard deviations. Could it be that Shine is the South African success equivalent to the Indian Shishuvachan classes?

Like the other Indian-based results in trained volunteer interventions as evaluated in Banerjee et al. (2007) and Lakshminarayana et al. (2013), Shine Literacy can attest to the potential that holds in recruiting an army of volunteers. Similarly to Banerjee et al. (2007), this analysis has shown that the lowest performers have benefitted the greatest from attending Shine. One of its most laudable results is that Shine Literacy aids those at the bottom the most: only 12.6% of those who were considered 'At risk' at baseline stayed in that category by endline. In fact, 23.7% moved up to 'Satisfactory' and 18.3% to 'Good' literacy. In the difference-in-differences results, a programme impact of 20.8% for 'At risk' learners was detected. In the same vein, only 9% of those who scored 'Poor' before the programme were still considered 'Poor' afterwards. The remaining fractions either moved up to 'Satisfactory' (37.5%) or – the large majority (53.3%) – increased by two literacy rankings to 'Good'. In contrast, those who were considered 'At risk' but were not assigned to Shine stayed there 81% of the time.

In effect, this means that Shine is very efficient in its focus on the learners that need attention the most. As opposed to education programmes that serve only high-achievers, such as information campaigns (Jensen 2010) or the provision of textbooks (Glewwe et al. 2009), Shine Literacy sets the focus on pulling those learners that struggle most out of a potentially never-ending literacy deficiency. The underlying goal is to put

14 Nevertheless, specifically with the difference-in-differences of -3% of 'Good' participants' test scores from Table 4 in mind, this movement should be observed more in-depth over time with more available data. What it could suggest, for instance, is that the programme is not well catered to those who are already doing well with regards to their literacy. It could be that the intervention is not fully fleshed out to accommodate those learners who are already at a good standing, and who would benefit more from regular class interaction than practicing literacy with a volunteer. Though it would imply sending volunteers away – which will rarely ever take place in practise – it could be that 'Good' learners might not benefit, and instead do worse. Thus, it may be worthwhile to explore this possibility and consider asking 'extra' volunteers to attend a different Shine Centre.

15 A thorough review of these is presented in the thesis on which this paper is based.

them on good footing for later school success. Most importantly, enabling learners' literacy to meet what is expected from them later by Grade 3 and Grade 4 ensures that they can proceed on their learning journey, and do not get left behind early on. In effect, learners do not need to exhaust their working memory becoming frustrated in trying to learn how to read, but can now read to learn.

Moreover, Shine Literacy does not only ensure that bottom-end learners increase their mean performance, but also that the distribution of test scores across the learner population becomes much tighter. The shifts in test score distribution before and after the intervention are immense, moving the combined mean of participants from 44.4% at baseline, to 66.4% at endline. Though non-Shine learners do better in literacy on average, those learners that were far away from such literacy levels before they attended the programme are coming closer to similar literacy proficiency afterwards. In practice, this means that instruction in the classroom can be streamlined more easily, thereby making the lives of already overstretched teachers much easier. Teachers are enabled to focus on the content that they are teaching, instead of trying to make sure that every learner understands a written text or instructions. This becomes particularly meaningful when considering that just employing more teachers and thereby changing pupil-teacher ratios may not actually have any effect on learners' performance, as indicated by the null (or negative) results found in Chin (2005), Urquiola (2006) and Duflo et al. (2015). In contrast, taking learners that seem to struggle with their reading literacy out of the classroom and giving them the opportunity to work on their individual shortfalls, one-on-one with a dedicated volunteer, proves to be effective in really improving performance.

Those interested in examining gender equity effects in schooling outcomes will be unsurprised that boy learners were more likely to be programme participants due to their lower literacy performance. However, they will also note that it seemed that participating boys benefitted just as much as girls. That is, Shine takes a step towards closing the achievement gap[16] between girls and boys when it comes to literacy. Moreover, bearing in mind that all schools examined in this analysis use only English as the language of teaching and learning, the immense literacy gains undergone by isiXhosa speakers are also indicative of Shine's success. Those learners who switch between speaking isiXhosa at home to speaking English at school experienced literacy improvements of 24.1%. This also implies that Shine successfully supports learners that do not speak English as their first language and also improves their English language proficiency – a large, and important stepping-stone to better English reading ability (Pretorius & Spaull 2016).

3.1 Keep shining bright: future research avenues and limitations

As a quasi-experimental impact evaluation of an education programme, the current analysis aimed to determine if there are empirical reasons to ascribe causality between Shine Literacy and learners' educational success. The effectiveness of Shine's support via trained volunteers is documented in this paper. However, it is important to remember that assignment to Shine Literacy is not randomised – and, at this point, it would be unethical to do so. As with many educational programmes that are evaluated in quasi-experimental studies, there are various factors that unravel in the real world that econometric methods simply cannot account for. In the Shine case, difference-in-differences estimates cannot hold time-variant factors constant. Similarly, not much is known about the learner population's innate ability, the factors they are exposed to at home, or even an indicator of how vested their parents are in their children's literacy. From Heckman (2006) and Heckman & Masterov (2004), it is known how important parental efficacy is in learners' academic development. Therefore, accounting for parental investment would be an interesting avenue to explore. Nevertheless, the difference-in-differences pointed to a strong and positive effect of

16 The reader interested in an assessment on the gender achievement gaps for both mathematics and reading is referred to Robinson & Lubienski (2011). Here, it is explored to what extend the reading gaps between boys and girls exist in the US context, and how this gap widens specifically among those at the bottom end of the performance distribution.

Shine Literacy. In the context of potential scale up, these effect sizes are educationally speaking very meaningful.

Several future avenues for evaluative work on Shine Literacy can be imagined. Especially with the similarity in success to the Shishuvachan classes in mind, it could be interesting to see how Shine Centres compare to their skeleton model counterparts: the Shine Chapters. These are set up as independently operated community franchises that deliver the Shine method. It is Shine's own way of scaling-up their model and making it more accessible and inexpensive to more parts of South Africa (Shine Literacy 2016). Currently, there are chapters in the Western Cape, the Eastern Cape and KwaZulu-Natal. It would be interesting to investigate if the literacy gains are similar to the Shine Centres' success story. In so doing, it could then also be compared in terms of cost-effectiveness. Being able to account for how much a standard deviation improvement actually costs would help when it comes to deciding on the scale-up of Shine across South Africa, and receive the dedicated support. Based on anecdotal evidence, one could assume that Shine does not need a lot more other than a specified learning space, some books and a dedicated volunteer.

There are additional interesting aspects to explore in future work. By itself, children's interaction with the enthusiastic adult volunteers has been identified as a great driver of learners' confidence and self-esteem. As put by Lufefe:

> I was too shy to speak. My Shine volunteer helped me to open up. It was scary at first, but then it was good. It's nice to know there are people who care about you. Leigh-Anne was my volunteer. She was wonderful. (Shine Literacy 2016: 12)

Judging by a large amount of anecdotal evidence recounted by the Shine staff and drawn on in their annual reports, participant learners have benefitted immensely by having dedicated, focussed attention given to them. To turn these testimonials into hard evidence, a behavioural experiment could be designed where learners would be given a small game to play before and after interacting with their volunteer. Increased confidence could be measured by asking children to bet on their future success in that game, similar to the method used by Hoff & Pandey (2014) who established a way to put an index on learner confidence. Of course, given that there it was slightly older learners participating, this experiment would have to be adapted to the early primary school case.

Similarly, given that Shine relies largely on its roster of volunteers, it would be interesting to explore what drives individuals to volunteer for Shine. By identifying which drivers function as motivators, it might become easier to target potential volunteers and grow Shine's volunteer roster. Lastly, it would be interesting to see whether the impact on mathematics persists beyond Grade 3.

Whilst these are all useful expansions, this paper provides sound evidence on the effectiveness of Shine in assisting learners that were deficient with their literacy. At baseline, Shine learners were a universe away from the rest of their cohort; at endline, this gap was greatly reduced. Perhaps, as Shine Literacy puts it, this is how words can change worlds.

REFERENCES

Angrist, J., Bettinger, E. and Kremer, M. (2006). 'Long-Term Educational Consequences of Secondary School Vouchers Evidence from Administrative Records in Colombia'. *The American Economic Review*, 96 (3): 847–862.

Baird, S., Mcintosh, C. and Ozler, B. (2011). 'Cash Or Condition ? Evidence From A Cash Transfer Experiment'. *The Quarterly Journal of Economics*, 126(4): 1709–1753.

Banerjee, A. V, Cole, S., Duflo, E. and Linden, L.L. (2007). 'Remedying Education: Evidence from Two Randomized Experiments in India'. *The Quarterly Journal of Economics*, 122 (3): 1235–1264.

Brooks, G., Miles, J., Torgerson, C.J. and Torgerson, D.J. (2006). 'Is an intervention using computer software effective in literacy learning? A randomised controlled trial'. *Educational Studies*, 32 (2): 133–143.

Chin, A. (2005). 'Can redistributing teachers across schools raise educational attainment? Evidence from Operation Blackboard in India'. *Journal of Development Economics*, 78 (2): 384–405.

Cohen, J. (1992). 'A Power Primer'. *Psychological Bulletin*, 112 (1): 155–159.

Department of Basic Education. (2016). *Education Statistics in South Africa*. Pretoria: Department of Basic Education.

Dhaliwal, I., Duflo, E., Glennerster, R. and Tulloch, C. (2012). 'Comparative Cost-Effectiveness Analysis to Inform Policy in Developing Countries'. *A General Framework with Applications for Education*. JPAL Working Paper December 2012. Boston, MA: MIT.

Duflo, E., Dupas, P. and Kremer, M. (2015). 'School governance, teacher incentives, and pupil-teacher ratios: Experimental evidence from Kenyan primary schools'. *Journal of Public Economics*, 123 (1): 92–110.

Elliott, J., Arthurs, J. and Williams, R. (2000). 'Volunteer Support in the Primary Classroom: the long-term impact of one initiative upon children's reading performance'. *British Educational Research Journal*, 26 (2): 227–244.

Gertler, P.J., Patrinos, H.A. and Rubio-Codina, M. (2012). 'Empowering parents to improve education: Evidence from rural Mexico'. *Journal of Development Economics*, 99 (1): 68–79.

Glewwe, P., Kremer, M. and Moulin, S. (2009). 'Many Children Left Behind ? Textbooks and Test Scores in Kenya'. *American Economic Journal: Applied Economics*, 1 (1): 112–135.

Glewwe, P., Ilias, N. and Kremer, M. (2010). 'Teacher Incentives'. *American Economic Journal: Applied Economics*, 2 (3): 205–227.

Grigg, D., Joffe, J., Okeyo, A., Schkolne, D., van der Merwe, N., Zuma, M., Mulenga, C., Boodhoo, A. (2016). 'The role of monitoring and evaluation in six South African reading programmes'. *Southern African Linguistics and Applied Language Studies*, 34 (4): 359–370.

He, F., Linden, L.L. and MacLeod, M. (2009). A better way to teach children to read? Evidence from a randomized controlled trial. JPAL Working Paper: May 2009. Boston, MA: MIT.

Heckman, J.J. (2006). 'Skill Formation and the Economics of Investing in Disadvantaged Children'. *Science*, June: 1900–1902.

Heckman, J.J. (2011). The economics of inequality: The value of early childhood education. American Educator: 31–36.

Heckman, J.J. and Masterov, D. V. (2004). The Productivity Argument for Investing in Young Children. Working Paper 5, Invest in Kids Working Group Committee for Economic Development.

Ho, A.D. and Yu, C.C. (2015). 'Descriptive Statistics for Modern Test Score Distributions: Skewness, Kurtosis, Discreteness, and Ceiling Effects'. *Educational and Psychological Measurement*, 75 (3): 365–388.

Hoff, K. and Pandey, P. (2014). 'Making up people - The effect of identity on performance in a modernizing society'. *Journal of Development Economics*, 106 (1): 118–131.

Howie, S., Venter, E. and van Staden, S. (2008). The effect of multilingual policies on performance and progression in reading literacy in South African primary schools. *Educational Research and Evaluation*, 14 (6): 551–560.

Jensen, R. (2010). 'The (Perceived) Returns to Education and the Demand for Schooling'. *The Quarterly Journal of Economics*, 125 (2): 515–548.

Justice, L.M., Kaderavek, J.N., Fan, X., Sofka, A. and Hunt, A. (2009). Accelerating Preschoolers' Early Literacy Development Through Classroom-Based Teacher-Child Storybook Reading and Explicit Print Referencing'. *Language, Speech, and Hearing Services in Schools*, 40 (1): 67–86.

Khandker, S., B. Koolwal, G. and Samad, H. (2010). *Handbook on Impact Evaluation: Quantitative Methods and Practices*. 1st ed. Washington, DC: The World Bank.

Lakshminarayana, R., Eble, A., Bhakta, P., Frost, C., Boone, P., Elbourne, D. and Mann, V. (2013). 'The Support to Rural India's Public Education System (STRIPES) Trial: A Cluster Randomised Controlled Trial of Supplementary Teaching, Learning Material and Material Support'. *PLoS ONE*, 8 (7): 1 - 17.

Lance, P., Guilkey, D., Hattori, A. and Angeles, G. (2014). *How Do We Know If a Program Made a Difference? A Guide to Statistical Methods for Program Impact Evaluation*. 1st ed. Chapel Hill, North Carolina: MEASURE Evaluation, University of North Carolina

Linnakylä, P., Malin, A. and Taube, K. (2004). 'Factors behind low reading literacy achievement'. *Scandinavian Journal of Educational Research*, 48 (3): 231–249.

Moloi, M.Q. and Chetty, M. (2011). Progress in Gender Equality in Education: South Africa. SACMEQ Policy Brief Number 6 (September). Paris: The Southern and Eastern Africa Consortium for Monitoring Educational Quality.

Mullis, I.V.S., Martin, M.O., Foy, P. and Hooper, M. (2017). Progress in International Reading Literacy Study 2016: International Results in Reading. Boston, MA: International Study Center, Boston College.

Overett, J. and Donald, D. (1998). 'Paired reading: effects of a parent involvement programme in a disadvantaged community in South Africa'. *British Journal of Educational Psychology*, 68(3): 347–356.

Du Plessis, E., Naude, H. and Viljoen, J. (2003). 'A conceptual framework for accelerating emergent literacy skills of disadvantaged pre-schoolers in South Africa'. *Educare*, 32 (1): 20–35.

Porter, C.L. and Johnson, J.E. (2004). 'Parents as Classroom Volunteers and Kindergarten Students' Emergent Reading Skills'. *The Journal of Educational Research*, 97 (5): 235–247.

Pretorius, E.J. and Spaull, N. (2016). 'Exploring relationships between oral reading fluency and reading comprehension amongst English second language readers in South Africa'. *Reading and Writing*, 29 (7): 1449–1471.

Robinson, J.P. and Lubienski, S.T. (2011). 'The Development of Gender Achievement Gaps in Mathematics and Reading During Elementary and Middle School: Examining Direct Cognitive Assessments and Teacher Ratings'. *American Educational Research Journal*, 48 (2): 268–302.

Salisbury, T. (2016). 'Education and inequality in South Africa: Returns to schooling in the post-apartheid era'. *International Journal of Educational Development*, 46 (1): 43–52.

Schkolne, D.S. (2014). An Outcome Evaluation of the Shine Centre's Literacy Hour Programme. Masters Thesis. Cape Town: School of Management Studies, University of Cape Town.

Shine Literacy. 2016. Shine Literacy Annual Report 2016. Wynberg: Shine Literacy.

Slentz, K.L., Early, D.M. and McKenna, M. (2008). A Guide to Assessment in Early Childhood: Infancy to Age Eight. Washington, DC: Washington State Office of Superintendent of Public Instruction.

Spaull, N. (2011). *Primary school performance in Botswana, Mozambique, Namibia, and South Africa*. SACMEQ Working Paper Number 8. Paris: The Southern and Eastern Africa Consortium for Monitoring Educational Quality.

The DG Murray Trust. (2012). The Shine Centre: Words can Change Worlds. Learning from our implementing partners. Learning Brief 11. Claremont: The DG Murray Trust.

The National Institute for Literacy. (2008). *Developing Early Literacy: A Scientific Synthesis of Early Literacy Development and Implications for Intervention*. Jessup, MD: The National Institute for Literacy.

Urquiola, M. (2006). 'Identifying Class Size Effects in Developing Countries: Evidence from Rural Bolivia'. *The Review of Economics and Statistics*, 88 (1): 171–177.

Wasik, B. A. (1998). 'Volunteer Tutoring Programs in Reading: A Review'. *Reading Research Quarterly*, 33 (3): 266–291.

About the author

This work is part of a thesis presented by Charlotte Stollberg for the degree of Master of Commerce (Applied Economics), School of Economics, University of Cape Town, Cape Town, South Africa. With experience in the NGO sphere, she has gathered some valuable experience in the set-up of randomised control trials, and programme evaluation in general.

Contact: Charlotte.Stollberg@alumni.uct.ac.za

Acknowledgments

This paper was produced within the framework of SA-TIED's Young Scholars Program. The program is designed to encourage promising university students to pursue careers in research.

INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE
A world free of hunger and malnutrition
1201 Eye Street, NW | Washington, DC 20005-3915 USA
T: +1.202.862.5600 | F: +1.202.862.5606
Email: ifpri@cgiar.org | www.ifpri.org