# AN OUTCOME EVALUATION OF THE SHINE CENTRE'S LITERACY HOUR PROGRAMME

DEBORAH STACEY SCHKOLNE

(GTZDEB001)

A dissertation submitted in partial fulfilment of the requirements for the award of the Degree of Master of Philosophy (Programme Evaluation)

University of Cape Town

School of Management Studies



## COMPULSORY DECLARATION:

This work has not been previously submitted in whole or in part for the award of any degree. It is my own work. Each significant contribution to and quotation in this dissertation from the work or works of other people has been attributed, cited and referenced.

Signature: Signed by candidate

Signature Removed

Date: 12 December 2014

## ACKNOWLEDGEMENTS

I would like to thank my supervisor, Joha Louw-Potgieter, for all her support, insights and patience throughout the year. Her engaging conversations and exacting standards helped me achieve a dissertation of which I am proud.

I would also like to thank Katya Mauff, my statistical consultant from the UCT Department of Statistical Science, for her valuable contribution to the results chapter.

Disclaimer: Any opinion, findings and conclusions or recommendations expressed in this document are those of the author(s) and the Statistical Consulting Service (UCT) does not accept responsibility for the statistical correctness of the research results reported.

## EXECUTIVE SUMMARY

This dissertation reports on an outcome evaluation of The Shine Centre's literacy intervention, Literacy Hour. The Shine Centre is a South African non-governmental organisation that offers literacy support to Grade 2 and 3 learners at risk of reading failure. The programme takes the form of one-to-one or one-to-two tutoring. The learners who are chosen to participate in the Literacy Hour programme are selected based on their results in Shine's diagnostic assessment completed at the end of Grade 1. The selected students work with trained volunteers twice a week for an hour at a time. Each hour of the Literacy Hour programme is divided into four 15-minute components, namely: (a) paired reading, (b) shared reading, (c) have-a-go writing, and (d) wordplay.

The evaluation design was a quasi-experimental, non-equivalent group design. The assignment into the experimental/intervention group and the control group was done by means of a sharp regression-discontinuity cut-off. The sample included six Centres and two Chapters in the Western Cape with complete data for the 2011 to 2013 cohort. Learner results were monitored using various assessment tools at the end of Grade 1 and during Grades 2 and 3. The two Shine diagnostic assessments, D1 and D2, tested the learners' literacy skills against the level that they should have acquired by that point in their school career. Additionally, the reading level assessments were used to determine the learners' actual reading level/age.

The first question this evaluation aimed to answer was whether the English literacy level of Grades 2 and 3 learners on the Literacy Hour programme increased to grade appropriate levels. The approach to this question was to compare the results of the intervention group to Shine's grade appropriate level on each assessment. The learners' progress in these assessments was also compared to the targets set by Shine.

The results illustrated that Shine's targets were generally met or exceeded. After only six months on the Shine programme most learners demonstrated literacy results at or above grade level. After one year on the programme most of the learners maintained their achievement in reaching grade level in both the diagnostic and reading assessment. After one year, 66% of the intervention learners graduated from the programme and only the weakest learners from the initial sample continued onto the second year. Although not all the learners in the sample reached

grade appropriate level in the reading test after one and a half years, the majority did reach grade appropriate level on the skills test after two years on the programme.

The second question focused specifically on learners near to the grade level cut-off point (77.77%). A regression discontinuity analysis was used to determine whether the learners allocated to the intervention group improved at a rate that was significantly greater than that of learners in the control group. I compared results for these two groups at the point of D1 assessment mid-Grade 2 and the reading level assessments at one year and one and a half years. Parametric and non-parametric models were tested to determine the best fit for the data.

The results demonstrated that after six months the improvement in the learner results was as a direct result of the literacy hour programme at a 5% significance level. After one year, although the difference in the results of the intervention and control group did not show significance at a 5% level, it did show significance at a 10% level. At the one and a half year reading assessment, I could not conclude that the improvement was a result of the programme at either the 5% or 10% significance level. This result resonates with the literature and was therefore not unexpected.

In conclusion, it is clear that Shine's Literacy Hour programme has a very successful foundation from which to work. While there are opportunities to improve certain elements, there is no doubting the fact that learners are measurably benefiting from the programme.

**TABLE OF CONTENTS**

**LIST OF FIGURES**

**LIST OF TABLES**

# INTRODUCTION

The aim of this dissertation is to evaluate the effectiveness of The Shine Centre's literacy intervention, Literacy Hour. The Shine Centre, hereafter referred to as Shine, is a South African non-governmental organisation (NGO) that offers literacy support to Grade 2 and 3 learners at risk of reading failure. Its reach extends to the Western Cape, Gauteng and KwaZulu-Natal provinces.

Founded on a desire to improve the literacy levels across the country, Shine envisions "a nation of readers" (http://www.theshinecentre.org.za). Its slogan, "words can change worlds", embodies a belief that "enabling a child to read enables the child to learn" (Weissenberg et al., 2014, p. 5).

It is important to understand the educational context in which this programme exists. I therefore report on the most recent national and international literacy reports. These include the Annual National Assessments (ANAs), the Progress in International Reading Literacy Study (PIRLS) and the Regional Southern and East Africa Consortium for Monitoring Education Quality report (SACMEQ).

In 2011, the South African Department of Basic Education (DBE) introduced ANAs for learners in Grades 1 to 6, as well as learners in Grade 9. These assessments, which focus on literacy and numeracy, act as a critical measure for monitoring the progress in learner achievement against the targets set out by the DBE (2013). Although the ANAs provide a standardised indication of learning, they have come under scrutiny because of their current implementation and lack of external verification and can therefore not be used as a reliable indicator of progress (Spaull, 2013).

The PIRLS compares the literacy results of Grades 4 and 5 learners from various countries using a standardised assessment (Howie, van Staden, Tshele, Dowse, & Zimmerman, 2011). The results are reported on a scale ranging from 0 to 1,000 points. The international reference point for literacy performance is a score of 500 points (Howie et al., 2011), which is set to correspond to the mean of the overall achievement distribution (Mullis, Martin, Foy, & Drucker, 2011). In the 2006 PIRLS South Africa scored the lowest out of the 45 participating

countries (Spaull, 2013). In 2011, the most recent PIRLS assessment, a new assessment called prePIRLS was introduced for countries whose performances in the 2006 study were deemed to be too low (Howie et al., 2011). Following its poor results in the 2006 PIRLS, South Africa participated in the 2011 prePIRLS. South Africa was one of three countries to participate in the prePIRLS assessment, including Botswana and Columbia (Howie et al., 2011). Learners were tested in their home language (any of the 11 official languages), yet 71% of the Grade 4 learners only reached the Low international benchmark (400 to 474 points). The average of 461 points was significantly lower than the prePIRLS centre point of 500 (Howie et al., 2011). Although South Africa scored similarly to Botswana, South African Grade 4 learners were, on average, 2.9 years below the average Columbian learner, even though South Africa's education expenditure is 49% higher than in Columbia (Spaull, 2014).

SACMEQ III is the most recent reported SACMEQ survey, conducted in 2007 (Spaull, 2011). It focuses on learners' schooling and home environments, testing Grade 6 learners' literacy skills in 15 participating Sub-Saharan countries (Spaull, 2012). The proportion of South African Grade 6 learners who were deemed functionally illiterate (unable to read a short, simple text and extract its meaning) was 27%. This compared poorly to the average of 18% for all participating countries. South Africa ranked tenth out of the 15 participating countries in mean reading score (Spaull, 2011). Over the period of seven years from SACMEQ II (2000) to SACMEQ III (2007), South African learners showed no improvement in Grade 6 literacy (Spaull, 2013).

These studies suggest that the scope of the education (and in particular, literacy) problem in South Africa is very large. Godwin Khosa (n.d.), CEO of The National Education Collaboration Trust, explains how a collaborative approach between government, business, teacher unions, NGOs, community, traditional and religious leadership, and parents is required to transform the quality of basic education in South Africa. It is therefore evident that the NGO sector, including literacy NGOs such a Shine, is expected to play a part in the solution.

In order to understand how Shine tackled this problem, a detailed description of its programme, the theory underlying the programme, a discussion around the plausibility of the theory, and evaluation questions based on the programme theory, are provided.

**Programme Description**

Shine's objective is to increase the number of children from disadvantaged backgrounds that complete Grade 3 with grade-appropriate literacy scores, independent learning habits, and academic self-confidence. The organisation strongly believes in creating a "thinking environment" (Weissenberg et al., 2014, p.7), a place in which each child has the time to formulate his or her own thoughts and ideas. Its primary intervention, a programme called Literacy Hour, is consistent with this ethos. It takes the form of one-to-one or one-to-two tutoring that targets learners at risk of reading failure in Grades 2 and 3. The learners who participate in the Literacy Hour programme are selected based on their results in Shine's diagnostic assessment completed at the end of Grade 1, and they work with trained volunteers twice a week, an hour at a time, for one or two years.

The first Shine Centre was established in South Africa in 2000. It is based on a model from the United Kingdom that encourages community members to assist with learner literacy support. The programme was piloted at Observatory Junior School in Cape Town, where struggling learners were paired with trained volunteers (Weissenberg et al., 2014). In the first year the implementers focused on Grade 5 learners; however, they realised that the programme needed to focus on earlier grades to be more effective in increasing learner's literacy results (M. Weissenberg & C. Mashek, personal communication, February 20, 2014). Through trial and error, Shine learnt many lessons regarding the target group, the structuring of sessions and the selection of learners (M. Weissenberg & C. Mashek, personal communication, February 20, 2014). Shine now has a product which they deem to be standardised and replicable, with all of its programme procedures documented in operational, volunteer and teacher manuals.

Shine has two distinct ownership models, namely Centres and Chapters. The Centres are run by Shine employees and are overseen by programme managers who report regularly to the Shine Head Office in Cape Town. By contrast, Shine Chapters are franchises whose employees are independent of Shine. The Chapters report to the Head Office bi-annually and are given support during their initial implementation and intermittently through the year, including training facilitated by Shine (M. Weissenberg and C. Mashek, personal communication, February 20, 2014). Shine's expansion and sustainability plan is to set up more Chapters, as this allows Shine

to extend its geographical reach without having to manage the programme closely with its own employees (Mashek, March 7, personal communication, 2014). Currently, there are eight Centres and five Chapters in Cape Town, four Chapters in Johannesburg and two Chapters in Kwazulu-Natal.

All of the Shine Centres are based at schools in which English is the language of instruction, despite the fact that English is not the first language for the majority of the learners. In contrast, some of the Chapters are based at schools in which the language of instruction is Zulu or Xhosa. Regardless, the Shine programme is presented in English (M. Weissenberg and C. Mashek, personal communication, February 20, 2014).

Shine requires volunteers to attend two initial training sessions prior to working with any learners. Thereafter, the volunteers are expected to attend another four training sessions over the course of the year, although these sessions are not compulsory. In addition to training, the volunteers receive on-site observation and feedback (M. Weissenberg and C. Mashek, personal communication, February 20, 2014)

In order to illustrate how learners progress through the Shine programme, a service utilisation plan was developed. This is shown in Figure 1. According to Rossi, Lipsey and Freeman (2004) a service utilisation plan is the journey of a learner through the programme from entry to exit, and it outlines the programme's assumptions and expectations about how to reach the target population and provide and sequence service contacts.

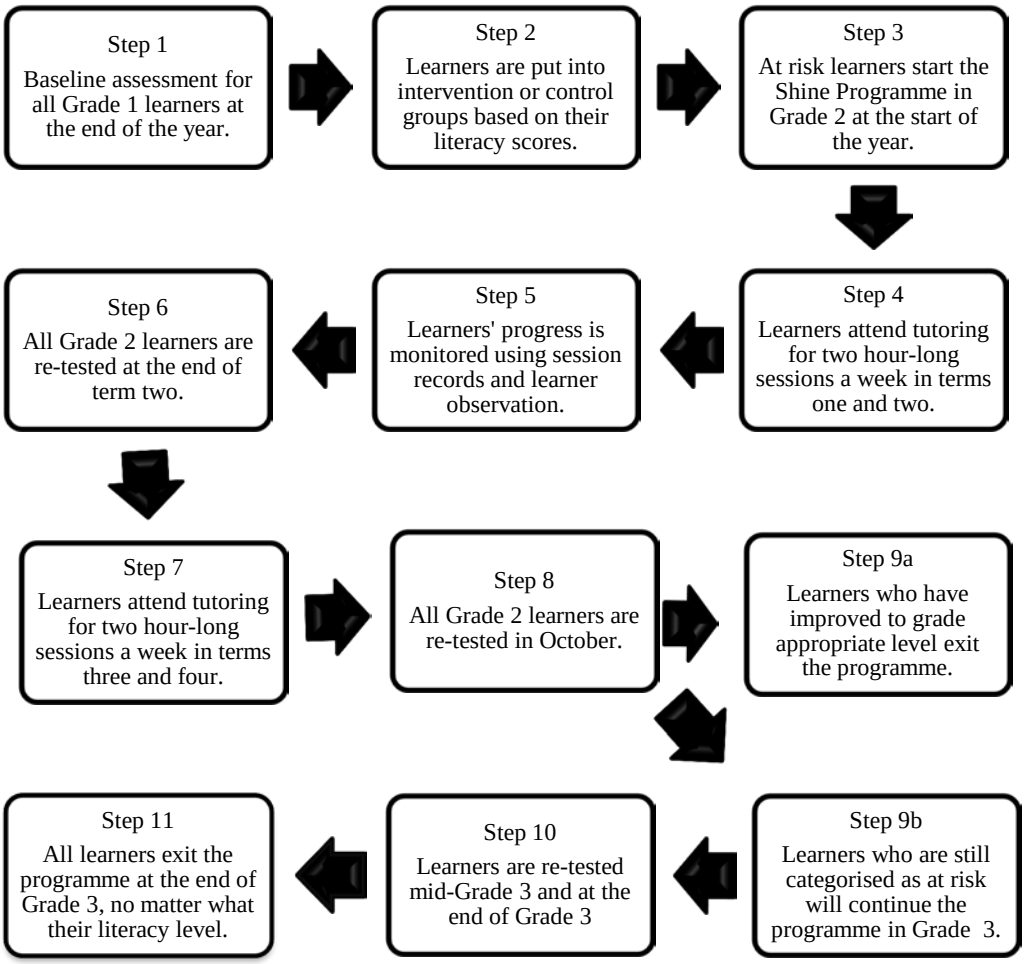


*Figure 1.* Shine Literacy Hour service utilisation by beneficiaries

Each hour of the Literacy Hour programme is divided into four 15-minute components. These components are outlined below as they have been described in the Shine Learning Partners Training Manual (Weissenberg et al., 2014).

**Paired reading.** In this component, the learner and the tutor read together a book that is matched to the learner's reading level. According to Weissenberg et al. (2014) this technique improves the learner's knowledge of basic rules of language and sentence construction, for example word spacing and punctuation marks; comprehension skills, including inference and predictions; memory of sight words; letter-sound knowledge and reading fluency and expression.

**Shared reading.** In this component the learner chooses a book for the tutor to read out loud to the learner. The tutor encourages comprehension through asking questions and discussion. According to Weissenberg et al. (2014) this technique increases the learner's enjoyment of books as well as giving the learner exposure to books, themes, and words beyond the scope of their own reading. It also improves imaginative thinking, comprehension skills, vocabulary, and self-esteem while decreasing the learner's anxiety.

**Have-a-go writing.** In this component the learner is encouraged to write a sentence using lines, written by the volunteer, as prompts to represent a word. According to Weissenberg et al. (2014) this technique improves the learner's knowledge of sentence construction, sentence writing skills, paragraph writing skills and sequencing sentence writing, while also increasing vocabulary, self-esteem, and general knowledge.

**Word play.** In this component the tutor uses specially-designed Shine Centre Literacy Games to teach the learner the necessary literacy skills required when learning to read. Examples include asking the learner to sound out words and blend sounds in order to make words. According to Weissenberg et al. (2014) this technique improves the learner's letter-sound skill, sounding out and blending sounds to form words, reading and writing words, reading common sight words, and ability to answer comprehension questions.

**Praise notes.** At the end of each Literacy Hour session, the tutor writes a praise note for the learner to take home. This note takes the form of a simple sentence acknowledging something positive about the learner and his or her progress. According to Weissenberg et al. (2014) this technique increases the learner's self-esteem, confidence and motivation. Prior to

leaving, the learner chooses a book from the Centre or the Chapter to take home for a few days to read.

The Shine Centres and Chapters provide schools with several services in addition to the Shine Literacy Hour Programme. They equip Grade R, Grade 1 and Grade 2 classes with a range of multicultural storybooks to foster curiosity about books and a love of reading. They run workshops annually for parents of Grade R and Grade 1 learners that focus on the importance of paired reading, reading stories to children and ensuring that their child's reading journey is filled with delight, praise, and support. They also meet with the teachers and principals to discuss the learners' progress and share insights and good practices. Lastly, they facilitate eye testing of all learners in Grade 1 and those learners who are new to the school, and they provide spectacles where necessary (M. Weissenberg and C. Mashek, personal communication, February 20, 2014).

**Programme Theory**

Programme theory is an underlying element in every social programme that depicts how the programme intends to create social benefit. According to Rossi et al. (2004) the programme theory is the logic that connects the programme activities and its intended outcomes (i.e. how the programme expects to solve the problem). The plausibility of the programme theory is essential because if it is faulty or if it is valid but not operationalised then the intended social benefits will not be achieved (Rossi et al., 2004).

Shine's programme theory was obtained by reviewing the Training Manual for Shine Partners (Weissenberg et al., 2014), visiting one of the Shine centres as well as conducting interviews (Weissenberg and Mashek, personal communication, March 19, 2014). The main assumption underlying this theory is that a Grade 2 or Grade 3 learner at risk of literacy failure who receives the Literacy Hour programme will improve his or her literacy to grade appropriate level.

Shine's programme theory is depicted in Figure 2.

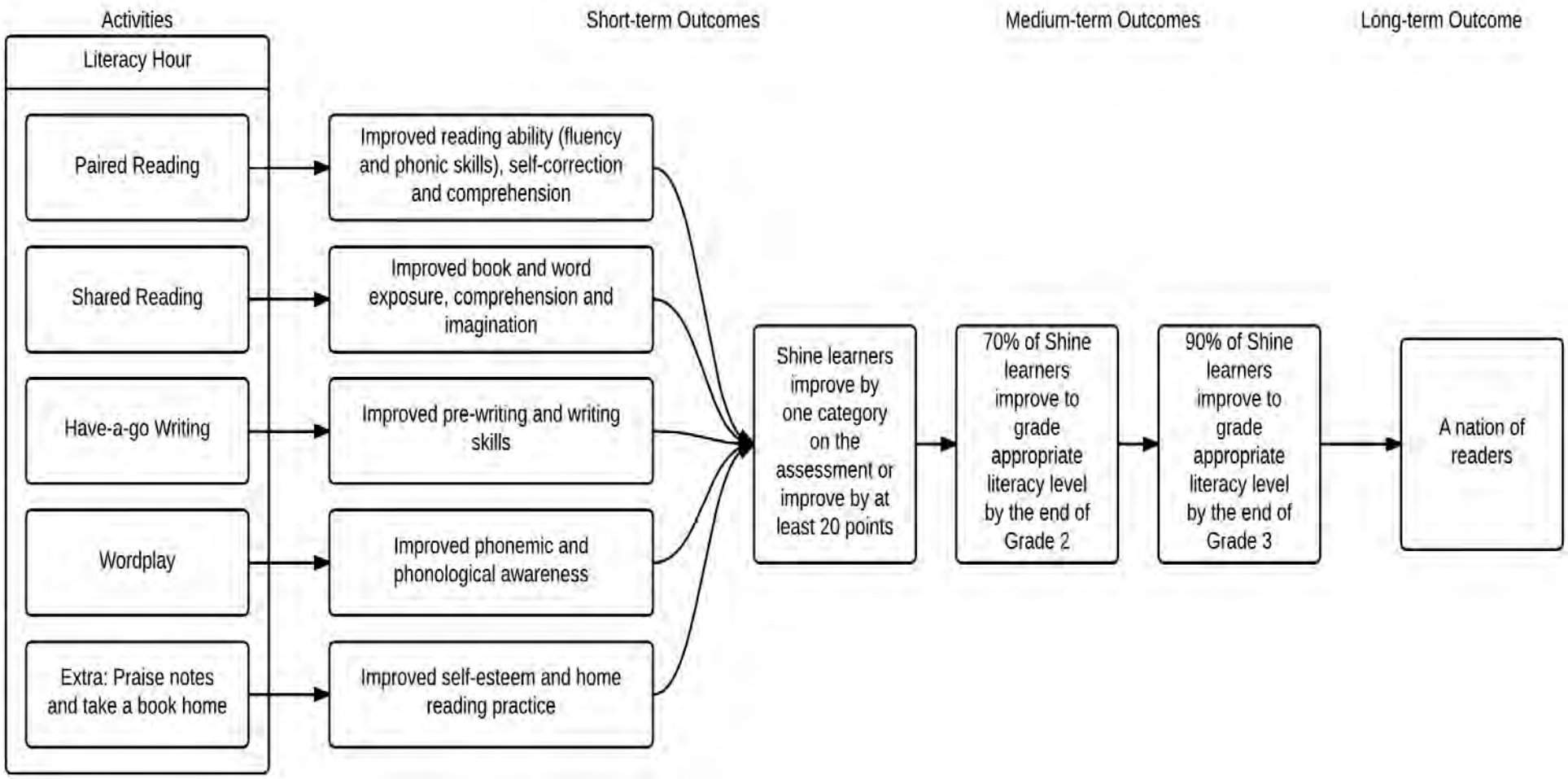


*Figure 2.* Shine programme theory

From Figure 2 one can conclude that if Grade 2 learners at risk of literacy failure attend the Literacy Hour programme, their literacy levels should improve to grade appropriate level by the end of Grade 3.

**Plausibility of Programme Theory**

The literature review that follows assesses the plausibility of Shine's programme theory. The review attempts to assess whether a tutoring programme can be effective in increasing literacy rates of South African learners at risk of literacy failure. I start by reviewing each of the components of the Literacy Hour programme against the literature. I then discuss what the literature suggests are the most effective implementation and target strategies to achieve superior literacy results taking into account the South African context.

In order to find relevant literature, a search was performed using social science databases, EBSCO, Google Scholar and Jstor. The search involved the use of specific key words, for example paired reading, shared reading, writing scaffolding, learning to write using scaffolding, learning through literacy games, primary school literacy interventions, primary school literacy interventions for second language learners, paired reading primary school, shared reading literacy interventions, evaluation AND literacy, tutoring AND literacy intervention, reading intervention, literacy AND South Africa, reading AND second language AND elementary, South Africa AND second language, PIRLS, SACMEQ, and Annual National Assessments.

The first component of the Literacy Hour programme is Paired Reading. While the literature refers to paired reading programmes using parents rather than tutors, it should be noted that both parents and tutors play a mentoring role in paired reading. Consequently, the interventions in the literature are still relevant to Shine.

Cadieux and Boudreault (2005) describe paired reading as being effective even with parents who lack high levels of education or who have poor reading abilities. This method comprises two phases. To begin, the learner and the parent simultaneously read a text chosen by the learner. This phase is important because it affords the learner the opportunity to model his or her reading on that of a fluent reader (Cadieux & Boudreault, 2005). Once the learner feels

ready, he or she may then signal to read independently. When the learner makes a mistake, the parent is on hand to make a correction within a few seconds. Immediately thereafter, both the learner and the parent will resume simultaneous reading until the learner signals that he or she is once again ready to read independently. Topping (1995) describes several advantages of this technique, including the learner being in control, the learner receiving lots of praise, a strong emphasis on understanding and a model to help pronounce difficult words. This technique should also lead to greater fluency, use of context, phonic skills, and likelihood of self-correction.

Cadieux and Boudreault (2005) evaluated the effectiveness of this method in a pre- and post-test. Both the control and intervention groups comprised Grade 1 learners and their parents. Both groups were given materials but only the intervention group was given training on the paired reading technique. The results reflected that the intervention learners made significant gains in phonetic awareness, but not in reading. The authors explained that this result is appropriate given that the learners were at the pre-reading level. Many of the Grade 2 learners entering the Shine programme have poor reading skills and may be at the pre-reading level. Cadieux and Boudreault's (2005) results demonstrate that paired reading can assist in building foundational skills even for learners such as these. In light of the literature, it seems plausible that including paired reading in the Literacy Hour programme could help to increase learners' phonetic awareness.

Rasinski, Fredericks, and Rasinski (1991) also evaluated a paired reading programme. The parents were trained and contracted to do paired reading with their learners for five minutes a day for three months. The results illustrated that learners became more fluent in reading. The teachers observed less word-by-word reading and noted sight word recognition and improvements in vocabulary.

The second component of the Literacy Hour programme is Shared Reading. This technique involves a tutor, parent or teacher reading aloud to the learner or class. An important question arising in the literature is whether reading aloud to learners in isolation will promote reading skills development, such as understanding of print, or whether one needs to focus directly on these elements to achieve results (Phillips, Norris. & Anderson, 2008; Piasta, Justice, McGinty, & Kaderavek, 2012; Zucker, Cabell, Justice, Pentimonti, & Kaderavek, 2013). Examples of print knowledge and literacy skills include knowing about the author and illustrator,

knowing that one reads from left to right, knowing upper and lower case letters and being able to make comparisons and predictions (Piasta et al., 2012; Zucker et al., 2013). Although reading books aloud may be interesting to children and may facilitate oral language development, Phillips et al. (2008) contend that this in itself does not foster emergent literacy development. The authors argue that supplementing reading aloud with explicit teaching about print has proven benefits for reading ability. This argument has been echoed in two longitudinal studies (Piasta et al., 2012; Zucker et al., 2013), which demonstrated that print knowledge and inferential talk are directly linked to children's reading outcomes. Using a randomised control trial (RCT) design, Piasta et al. (2012) assessed the effect of increasing pre-schoolers' attention to print during reading. The intervention classrooms experienced shared reading two to four times a week with the teachers verbally and nonverbally referencing print, whilst the control classrooms experienced the teachers' typical book reading style. The results revealed that the use of print references had significant effects on children's early literacy skills (e.g., reading, spelling, and comprehension) for two years following the completion of the RCT, demonstrating that print knowledge has a causal relationship with literacy skills and has important implications for reading development (Piasta et al., 2012). Zucker et al. (2013) examined the longitudinal relationship between the frequency and features of reading experiences within a pre-school classroom and the children's language and literacy outcomes in Kindergarten and Grade 1. This extra-textual talk during pre-school had significant positive results extending to Grade 1 skills Zucker et al. (2013).

The literature suggests that while reading aloud to the Shine learners could plausibly increase their imagination and exposure to books, this technique will only have a significant effect on early literacy skills if print knowledge and inferential talk are also explicitly targeted. Since Shine's tutors do use extra-textual talk in this component of the Literacy Hour, it seems plausible that the Shine learners will develop print knowledge which should lead to increased literacy and reading skills.

The third element of the Literacy Hour is Have-a-go Writing. This is a scaffolding method which facilitates emergent writing and assists in the transition to independent writing (Bodrova & Leong, 1998). The authors explain that a line is used to represent each spoken word. The learner creates an oral sentence and thereafter the teacher, parent or tutor helps the learner to

draw a line to stand for each word in the sentence. Thereafter, the learner tries to fill out the empty line spaces. Depending on the level of the learner, this could take the form of scribbles, letter-like forms or actual words. Eventually, the child discontinues the use of the lines altogether (Bodrova & Leong, 1998).

The studies by Bodrova and Leong (1998) and Graham, McKeown, Kiuhara and Harris (2012) demonstrate the effectiveness of this technique. Bodrova and Leong (1998) conducted a study with five year old learners from a low-income, multi-ethnic urban school. The students at this school were considered to be an at-risk population. The scaffolding writing technique produced more advanced writing compared to the level of writing the children produced when not assisted (Bodrova & Leong, 1998). This was demonstrated by the intervention learners' using more advanced appearing forms of writing and increased length and quality of sentences. Graham et al. (2012) performed a meta-analysis of writing interventions in elementary grades (i.e., Grades 2 to 6). The eight studies that tested the effectiveness of pre-writing activities, such as this scaffolding technique, found that learners improved their writing quality when using them. It therefore seems plausible that the Shine learners' pre-writing and writing skills will develop as a result of this technique.

The last component of the Literacy Hour programme is Wordplay. In this section, tutors engage learners with a number of games focusing on phonemic and phonological awareness. Stuart (1999) evaluated the effect of having a phoneme segmentation and skills programme, as compared to the more traditional approach of emphasising meaning. The programme focused on five-year-old inner-city learners who spoke English as a second language and initially had poor receptive vocabularies in English. The programme took place in a classroom setting over 12 weeks with the learners divided into an intervention group and a control group. The results showed that the intervention had a positive effect on the development of sound-letter skills as well as improving reading and writing skills. This result suggests that it is plausible that the wordplay technique will lead to increased literacy skills amongst Shine learners, including those learners who are disadvantaged and speak English as a second language.

We have established that it is plausible that each of the four components of the Literacy Hour programme could independently lead to improvements in literacy. The literature suggests however that a comprehensive mixed approach can have even greater gains for learners (Elbaum,

Vaughn, Tejero Hughes, & Watson Moody, 2000; Wasik & Slavin, 1993). All four effective programmes reviewed by Wasik (1998) comprised several components, namely rereading a familiar text or story; word analysis such as letters in words, sounds, segmentation, phonics and blends; writing and introducing new stories. In principle, therefore, it seems plausible that the combination of the four elements of the Literacy Hour programme could lead to greater improvements in Shine learners' literacy results.

With this in mind, it is instructive to review the structure and implementation of the programme. The initial focus areas are the tutoring frequency and duration, the relationship between the tutor and the learner and the approach to monitoring learners' results. I then discuss the choice of grades for the intervention, specific concerns around very weak learners and the choice of tutors. Lastly, I examine the South African context, with a focus on learners for whom English is a second language.

The effectiveness of the programme is influenced by the frequency and duration of the tutoring sessions. Wasik (1998) emphasises that tutoring needs to be intensive and consistent, for at least one and a half to two hours per week. In their meta-analysis, Elbaum et al. (2000) echo that it is most effective if the instructional time of the intervention is delivered intensely. The Literacy Hour programme, which comprises two hour-long sessions per week, would appear to have adequate time.

Wasik (1998) highlights the importance of relationship building and recommends that the learner should be tutored by the same person every week. This is the case for most learners on the Shine programme. While in rare cases learners might have two or three tutors in any given year, this still allows for relationship building because of the one-to-one or two-to-one ratio. This relationship building helps the learner become more comfortable with the volunteer and also helps the volunteer assess the learner's progress over time (Wasik, 1998).

The monitoring of the learners' progress is yet another element that is essential to the effectiveness of a tutoring programme. Wasik (1998) states that one needs to assess the learners on an ongoing basis in order for tutors to adapt to the learners' changing needs. Assessments should take the form of records after each session as well as formal assessments. The author also explains how the lessons being learnt from the assessments should form the basis for future

training of volunteers. The Shine Literacy Hour programme assesses the learners bi-annually. They also record the activities and progress of the learner after each session in a record book. This is consistent with the suggestions in the literature.

The literature also suggests that learners' grade levels are significantly associated with the variation in effect sizes of literacy interventions. In a meta-analysis of adult-instructed one-to-one reading interventions for elementary students at risk of reading failure, Elbaum et al. (2000) found the mean effect size to be moderate for learners in Grades 1, 2 and 3 but immaterial for learners in Grades 4 to 6. Therefore, literacy interventions need to focus on learners in early years. The Shine programme is consistent with this approach by focusing on learners in Grades 2 and 3.

Elbaum et al. (2000) raise a specific concern regarding learners with severe difficulties in reading. They note that for these learners the average effect size of the literacy intervention would likely be insufficient to raise performance to within the average range for a grade level. Further, these learners would likely need sustained support beyond the time frame of the intervention. Nevertheless, the researchers mention that the benefit of the intervention might yet be great enough to allow these learners to keep up with classroom instruction and to avoid academic failure. This concern is particularly relevant to Shine given that many learners on the programme have great difficulty reading. In the Literacy Hour evaluation questions, to be discussed later, Shine defines what it believes is a grade appropriate literacy level. The higher this bar is set, the greater is the likelihood that the effect size of the Literacy Hour programme will be insufficient for weak learners to reach this level.

The literature suggests that it is plausible that a volunteer, student or paraprofessional can be effective in implementing a tutoring programme. A meta-analysis of 96 studies concluded that teachers are more effective than teaching assistants and volunteers (Lake, Slavin, Susan, & Nancy, 2009). However, using teachers for tutoring interventions can be prohibitively expensive in South Africa given that most at-risk learners come from disadvantaged schools and areas.

Wasik (1998) acknowledges that while volunteers will never replace certified teachers, they can still be effective if the tutoring is founded on a number of guidelines, including having a certified reading specialist to supervise the tutors. This person needs to coordinate the

programme and must be knowledgeable about reading and problems in reading. In addition, tutors need ongoing, targeted training and feedback through on-site observation. The Shine Literacy Hour programme uses trained volunteers who receive regular training and feedback. Furthermore, a manager is on site for every session for quality control purposes. This is consistent with the suggestions in the literature.

No matter how well a programme is planned, it is likely to fail if it does not take into account contextual constraints. Consequently, it is important to consider the plausibility of the Literacy Hour programme in the South African context.

One of the main predictors of learner failure in South Africa is socio-economic status (Hungi, 2011). The author points to the results from the 2000 and 2007 SACMEQ reports indicating that learners from wealthy schools significantly outperformed their counterparts from poor schools. School-specific predictors include school resources (Hungi, 2011). Howie et al., (2011) note that almost half of the South African learners who participated in the 2011 PIRLS study attended schools without libraries. Out of a total of 1,000 points, learners from these schools achieved an average of 155 points less than learners at schools with well-resourced libraries. Shine takes these elements into account by providing resources to all the schools in which they work and also allowing learners to take books home to practise reading privately.

The other critical contextual element is the fact that most learners on the Shine programme do not speak English as a home language. Ehri, Dreyer, Flugman, and Gross (2007) evaluated a one-to-one tutoring programme called Reading Rescue that targeted language minority Grade 1 learners. In this programme the tutored learners demonstrated greater gains in reading and comprehension than the control groups. Begeny, Ross, Greene, Mitchell, and Whitehouse (2012) evaluated the HELPS programme targeting Latino English language learners in Grade 2 at rural public schools. The learners were tutored two to three mornings a week outside their classroom for ten minutes per session. The authors concluded that the experimental group outperformed the control groups in reading fluency and comprehension. These examples suggest that one-to-one tutoring programmes can be effective with second language learners.

Based on the literature reviewed, it seems plausible that Shine's Literacy Hour programme could improve learners' literacy scores. The literature supports the programme's

activities, structure and implementation as well as most of the contextual considerations. The main area which may need further consideration is whether the very weak learners might struggle to reach the grade appropriate reading level.

**Evaluation Questions**

This evaluation will be an outcome evaluation, evaluating the state of the target population or social condition that the programme is expected to have changed (Rossi et al., 2004). In order to assess this change, the following evaluation questions have been derived from the programme theory of the Shine programme:

1) Did the literacy scores for the children on the programme improve to grade appropriate levels and did Shine reach their targets?
2) Was there any significant difference in the average literacy score of the children who received the programme and those who did not?

# METHOD

## Research Design

The most reliable, valid, and appropriate evaluation design for the Literacy Hour programme was a quasi-experimental, non-equivalent group design. The assignment of participants into an experimental (intervention) and a control group was done by means of a sharp regression-discontinuity cut-off.

When using a regression discontinuity cut-off, the allocation of participants into the intervention and control groups is not done randomly, but is based on whether they score above or below the identified cut-off score (Shadish, Cook, & Campbell, 2002). The assignment variable, in Shine's case the Grade 1 first diagnostic (or D1) assessment, was measured before treatment. The cut-off was 77.77%. This design implies that the treatment effect of the intervention is fully determined by the rating variable and its cut-off (Jacob et al., 2012). The assignment variable assessed the need, meaning that the learners who needed the intervention were below the cut-off and those that do not need the intervention were above the cut-off. This cut-off was also appropriate because it helped overcome some of the ethical considerations that could have arisen when using a randomised control trial (RCT) design in this type of setting. Specifically, the Shine Centres were required to take on all the learners who were in need of the programme, subject to their having sufficient capacity to do so. If this was not a consideration then an RCT would have been a stronger design.

The regression discontinuity cut-off has several advantages. It controls for many of the common threats to validity that may arise when not doing a true experiment (e.g., selection bias). This occurs when learners are not allocated randomly to the intervention and control groups. This problem was overcome by selecting a variable upfront by which to allocate learners into the respective groups. This predetermined allocation strategy eliminated the uncertainty about how selection was performed (Rossi et al., 2004).

Perraillon (2013) and Lesik (2006) explain that with a regression-discontinuity cut-off, differences between candidates whose scores fall just above or just below the cut-off point can be treated as being random. This is referred to by Jacob et al. (2012) as local randomisation. Consequently, these candidates were considered identical apart from the exposure to treatment of

those falling into the intervention group. Therefore, statistically significant difference in the mean outcomes for these two groups subsequent to the intervention can be attributed to the treatment.

**Sampling and Participants**

The data used in this evaluation were collected by Shine from the schools in which the Literacy Hour programme operates. Since its inception, more than 2,600 learners have benefited from Shine's literacy interventions (http://www.theshinecentre.org.za, 2014). However, the data used for this evaluation only relate to a subset of the total learners. The reasons for this are outlined below.

The assessments used to gauge learners' literacy progress have gone through several iterations over the years. This made it difficult to credibly compare results for different cohorts of learners. Therefore, I started by considering only the data relating to the single cohort of learners that participated in the programme from 2011 to 2013. This was the cohort for which the most recent complete assessment data were available.

In order to reduce the potential effect of geographical differences between provinces, I only considered Shine Centres and Chapters from the Western Cape. While learners from 13 schools across the province have participated in the programme to date, some of these schools only joined the programme in 2013 and thus have incomplete data. I therefore only included the six Centres and two Chapters with complete data for the 2011to 2013 cohort.

The details for these schools are outlined in Table 1.

Table 1

*Details of Schools Included in the Final Evaluation Sample*

| Shine Status | Schools: Western Cape | No. Learners in sample (n = 174) | No. of libraries | Language of learning and teaching | Fee per learner per year | Quintile |
|---|---|---|---|---|---|---|
| Centre | School 1 | 9 | 1 | English | R7000-R7499 | 5 |
| Chapter | School 2 | 16 | 1 | Par: Eng/Afr | <R500 | 4 |
| Centre | School 3 | 51 | 1 | English | R2500-R2999 | 5 |
| Centre | School 4 | 43 | 1 | English | R500-R999 | 5 |
| Centre | School 5 | 20 | 1 | English | <R500 | 5 |
| Chapter | School 6 | 9 | 1 | English | <R500 | 4 |
| Centre | School 7 | 17 | 0 | English | R1000-R1499 | 5 |
| Centre | School 8 | 9 | 0 | English | R1000-R1499 | 5 |

It is important to note that some of the schools showed material differences in terms of language of learning, and fee per learner. All else being equal, it is reasonable to assume that these differences could lead to variations between schools in terms of the success of the intervention. However, due to the small sample size within each school, I could not investigate whether these school differences were statistically significant. While consistency between schools is preferable, it did not make sense to remove any more schools from the sample on this basis given the limited number of schools and learners remaining.

As discussed above, Shine allocated learners into the intervention and control groups on the basis of the learners' English literacy results in their Grade 1 diagnostic 1 (or D1) assessment. The test was scored out of 90 and Shine considered learners who scored a mark of 70 (i.e., 77.77%) or higher to be at grade-appropriate level. This test was appropriate because it used continuous scoring, a requirement for regression discontinuity (Schochet et al., 2010). On the basis of the learners' test scores, Shine allocated learners into five distinct categories. Learners in the three lowest categories, namely (a) At Risk ++, (b) At Risk, and (c) Poor, were considered to be below grade level and are thus included in the Literacy Hour intervention group. Learners in the Satisfactory and Good categories were considered to be at grade level. These learners are excluded from the intervention and were considered to be the control group. All learners, regardless of how Shine allocated them, received the usual literacy tuition that forms part of the school curriculum.

According to Lesik (2006) a regression-discontinuity cut-off requires perfect compliance with the allocation criteria to determine whether a learner should form part of the intervention or control group. However, in some instances this was not strictly adhered to. Some learners ($n$ = 94) scored below the cut-off point but did not receive the intervention, possibly due to resource constraints in terms of number of tutors available at their school. Out of these 94 learners, 75 were learners from the Chapters. In other cases ($n$ = 18), learners scored above the cut-off point but still received the intervention, possibly due to their school having extra capacity. There were also a handful of learners ($n$ = 5) above the cut-off point who received the intervention at the expense of learners below the cut-off point. This occurred for three learners at School 4, one learner at School 2 and one learner at School 5. There were also some learners for whom data were not captured properly, including missing data as well as inaccuracies with categorisation. For example, 14 learners were categorised as having graduated at the point of the initial assessment, which is impossible.

In each of these cases, the learners were considered to be misclassified. In line with the research design principles explained above, they were excluded from the sample.

In total, 188 of the 483 potential learners at the eight schools included in the sample were incorrectly classified. 121 students were correctly classified as never attending the programme and 174 students were correctly classified as participating. These learners made up the initial control group and intervention groups respectively at the start of Grade 2 in 2011.

**Measures**

Learners were initially assessed prior to Grade 2, with subsequent assessments during Grades 2 and 3. Table 2 outlines the assessment timetable.

Table 2

*Assessment Tools and Timelines for Intervention and Control Groups*

| | Grade 1 Oct-11 | Grade 2 May-12 | Grade 2 Oct-12 | Grade 3 May-13 | Grade 3 Oct-13 |
|---|---|---|---|---|---|
| Intervention | Diagnostic 1 | Diagnostic 1 | Diagnostic 2 Reading assessment | Reading assessment | Diagnostic 2 |
| Control | Diagnostic 1 | Diagnostic 1 | Reading assessment | Reading assessment | (None) |
| Abbreviations | GR1D1 | GR2D1 | GR2D2 and GR2 Reading | GR3 Reading | GR3D2 |

*Note:* See abbreviations per column

The two Shine diagnostic assessments, D1 and D2, test the learners' literacy skills against the level that they should have acquired by that point in their school career. The results of these assessments were reflected as scores out of a total or as a percentage. In contrast, the reading level assessments were used to determine the learners' actual reading level/age. The results of these assessments are reflected as reading age in months.

**Data Analysis**

The analysis aimed to address the two main questions of this evaluation.

The first question this evaluation aimed to answer is whether the English literacy level of Grades 2 and 3 learners on the Literacy Hour programme increased to grade appropriate levels. The approach to this question was to compare the results of the intervention group to Shine's grade appropriate level on the two D1 assessments (at baseline and six months), where the standard is 77.77%, the two D2 assessments (at one year and two years), where the standard is 75% and finally the two reading level assessments (at one year and one and a half years) where the standard is 94 months for the reading test at one year and 101 months for the reading test at one and a half years. The learners' progress in these assessments was also compared to the targets set by Shine. The targets illustrated in the programme theory are that after six months most learners should improve by 22%, after one year 70% of learners should improve to grade-appropriate level and two years 90% of learners should improve to grade-appropriate level.

The second question focused specifically on learners near to the cut-off point. A regression discontinuity analysis was used to determine whether the learners allocated to the intervention group improved at a rate that was significantly greater than that of learners in the control group. I compared results for these two groups at the point of D1 assessment mid-Grade 2 and the reading level assessments at one year and one and a half years. Parametric and non-parametric models were tested to determine the best fit for the data.

**Ethics**

A proposal for the evaluation was approved by the Commerce Faculty's Ethics in Research Committee.

## RESULTS

Before the results are presented, the variability in the sample size of the intervention group needs to be explained.

The number of learners in the control and intervention groups decreased over time. The reason for this is that at the end of both Grade 2 and Grade 3, learners have the opportunity to graduate from the programme if they reach grade appropriate literacy/reading levels. While some learners also graduated from the programme after six months and after one and a half years, these are not official exit points so this is not encouraged. Graduation can only occur when a learner's results reach grade-appropriate level in the most recent assessment.

The intervention group sample therefore only includes learners who remain on the programme up to the specific assessment point. By comparison, at each of these points the control group sample only includes learners who started in the control group. Thus, learners who graduate at various points are excluded from subsequent samples rather than being added to the control group. The details are illustrated in Figure 3.

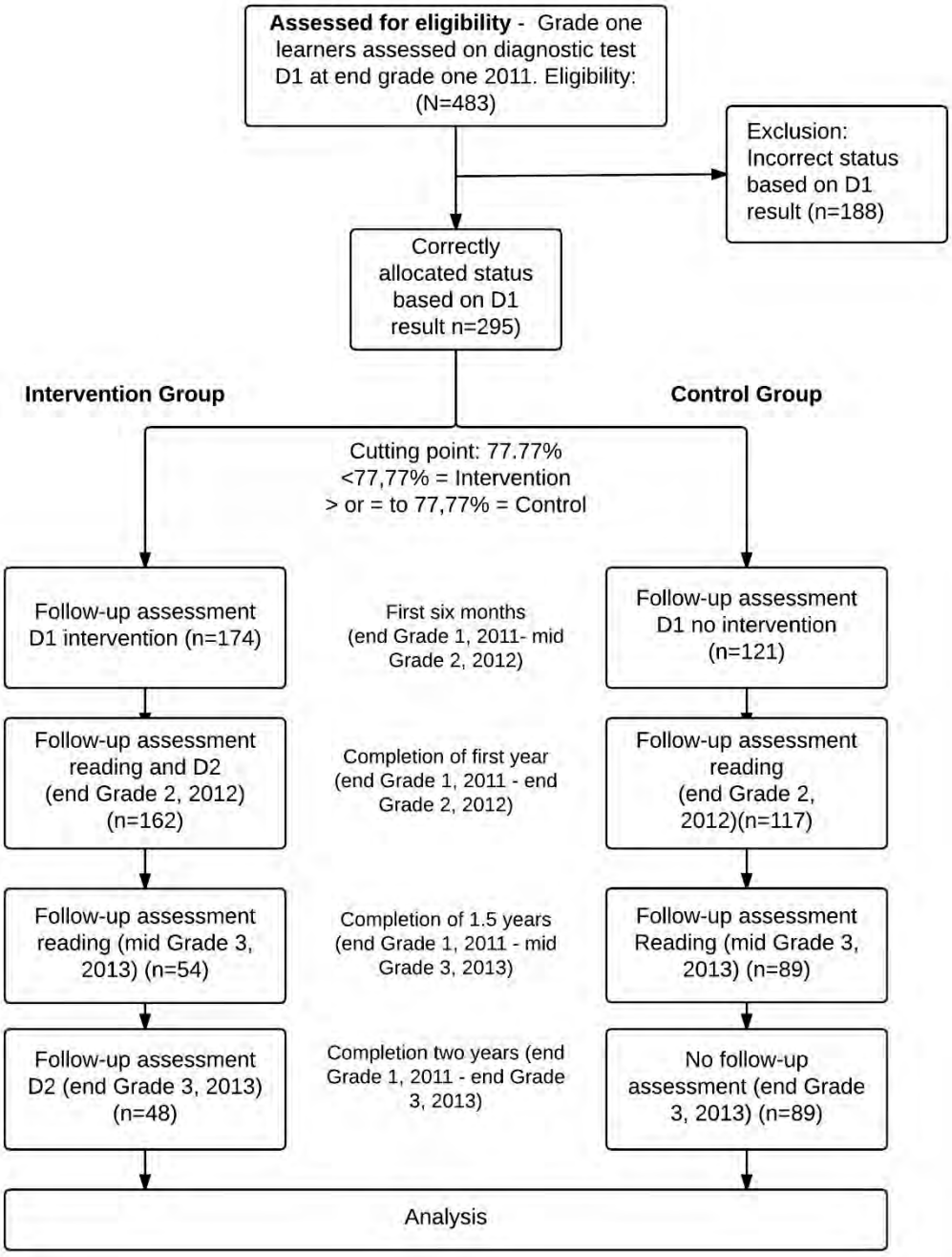


*Figure 3.* Number of learners in the sample at each of the assessment time points

Figure 3 clearly illustrates the sample changes throughout the two year timeframe. This sample will be used when determining the programme effect on learners at each of the time periods. However, when one analyses whether the Shine targets have been met, one needs to calculate the cumulative number of graduates at each time frame (i.e. all of the learners from the initial sample reaching grade appropriate level). This calculation was complicated due to various inconsistencies, illustrated in Table 3.

Table 3

*Sample Breakdown of Intervention Learners by Time on Programme and Results from Programme*

| School | Start (0) | GL on D1 (6) | Start (6) | GL on D2 (12) | GL on RA1 (12) | Start (12) | GL on RA2 (18) | Start (18) | GL on D2 (24) | Not GL (24) |
|---|---|---|---|---|---|---|---|---|---|---|
| School 1 | 9 | 9 | 7 | 7 | 6 | 6 | 6 | 6 | 5 | 1 |
| *School 2 | 16 | 3 | 12 | 0 | 1 | 9 | 2 | 6 | 1 | 5 |
| School 3 | 51 | 39 | 48 | 45 | 47 | 4 | 1 | 3 | 3 | 0 |
| School 4 | 43 | 29 | 42 | 28 | 29 | 19 | 6 | 19 | 16 | 3 |
| School 5 | 20 | 16 | 19 | 13 | 9 | 8 | 1 | 8 | 7 | 1 |
| *School 6 | 9 | 0 | 8 | 0 | 3 | 8 | 3 | 6 | 6 | 0 |
| School 7 | 17 | 12 | 17 | 13 | 14 | 0 | 0 | 0 | 0 | 0 |
| School 8 | 9 | 8 | 9 | 9 | 9 | 0 | 0 | 0 | 0 | 0 |
| Total | 174 | 116 | 162 | 115 | 118 | 54 | 19 | 48 | 38 | 10 |

*Note:* * Chapter.

| | |
|---|---|
| Start (0) | at inception of the intervention |
| GL on D1 (6) | reaching grade level in assessment D1 after six months |
| Start (6) | still on the programme after six months |
| GL on D2 (12) | reaching grade level in assessment D2 after 12 months |
| GL on RA1 (12) | reaching grade level in the first reading assessment after 12 months |
| Start (12) | still on the programme after 12 months |
| GL on RA2 (18) | reaching grade level in the second reading assessment after 18 months |
| Start (18) | still on the programme after 18 months |
| GL on D2 (24) | reaching grade level in assessment D2 after 24 months |
| Not GL (24) | not reaching grade level after 24 months |

It is important to look into the detail of Table 3, to understand the challenge in calculating the exact number of learners graduating from the programme.

At the start of the programme there were 174 learners in the intervention group. After six months on the programme, 116 of these learners (67%) reached grade appropriate level. Since this is not an official exit point of the programme, most learners do not leave the programme at this point. However, only 162 learners were still on the programme after six months. This means that 12 learners left the programme by this point for unknown reasons. It is possible that these learners left the school or for some or other reason missed the assessment. In rare cases, a learner may have left the programme because they had reached grade level and no longer required the intervention.

After one year on the programme, the learners were assessed using both the D2 assessment and the first reading assessment. Of the 162 learners who completed the full year on the programme, 115 reached grade appropriate level on the D2 assessment and 118 reached grade appropriate level on the reading assessment. In some cases, learners are kept on the programme if they do not reach grade appropriate level in both of these assessments. Since this is an official exit point one expects to see between 115 and 118 learners to leave the programme. However, only 54 out of the initial 174 learners started the programme in the second year. This means that 120 learners left the programme within the first year. This in turn means that between two to five learners left the programme for unknown reasons.

Of the 54 learners who were on the programme at the start of the second year, 19 learners reached grade appropriate level at the second reading assessment after one and a half years. Since this too is not an official exit point of the programme, one would not expect any learners to leave the programme at this point. However, only 48 learners were still on the programme after one and a half years. This implies that a further six learners dropped out of the programme for unknown reasons by this point.

Of the 48 learners still on the programme after one and a half years, 38 learners achieved grade level on the D2 assessment at the end of the second year. Therefore, after completion of the two year programme and all the assessments, 10 known learners of the initial 174 definitely did not reach grade appropriate level. In addition, there are between 20 and 23 learners that left

the programme over the two years for unknown reasons, which means that they may or may not have reached grade appropriate level. It is possible that these learners would have achieved grade level by the end of the second or year of the programme – I cannot know conclusively.

In some cases learners left the programme despite not reaching grade level. For example, at School 7 no learners started the second year on the programme despite not all of the learners' reaching grade level after one year on the programme. Conversely, in some cases learners remained on the programme despite reaching grade level. For example, at School 1 at least six out of the nine initial learners reached grade after one year on the programme, yet six learners started the second year on the programme.

These inconsistencies do not invalidate the results that follow. However, their presence means that I am unable to know the exact numbers of learners reaching grade level and graduating at any given point. This problem is exacerbated by the fact that some learners leave the programme for unknown reasons, as explained above. By ignoring all of these inconsistent learners, I am implicitly taking a conservative view of the total number of graduates, when comparing the total to the target set.

From this point onwards, the results will be presented in terms of the two evaluation questions.

**Evaluation question 1: Did the literacy scores for the children on the programme improve to grade appropriate levels and did Shine reach their targets?**

This question will be answered by considering each of different time periods that relate to when the learners in the intervention group were tested. At each assessment period, the results of the intervention learners are presented in light of Shine targets and grade level. All the assessment abbreviations can be found in Table 2, in the methods chapter.

The Shine targets are presented as the cumulative number of learners who reached the target, from the initial 174 learners who started the programme, at the specific assessment period. The calculation takes into account learners who reached grade level at the official exit points of one and two years and not learners who exited the programme at six months or one and a half years.

The results which compare learner results to grade appropriate level include only the learners who participated in the programme during that specific period (i.e. if learners exited the programme after one year, they were not included in the results at the two year assessment), as illustrated in Figure 3.

**October 2011 to May 2012 (6 months on programme)**. During this period, 130 out of 174 intervention learners improved by at least 20 points (22%) from the initial assessment to the second assessment. This total of 74.71% exceeds the target set by Shine, which aimed that 50% of the learners would increase by at least 20 points.

In the initial diagnostic assessment performed in October 2011, GR1D1, the intervention group achieved a mean score of 45.50% and a median score of 49.44%. In the second diagnostic assessment performed in May 2012, GR2D1, the intervention group achieved a mean score of 77.94% and a median score of 84.44%. This means that the average increase in learner mean and median scores was 32.44% and 35% respectively. These details are outlined in Table 4.

Table 4

*Six Month Measure of Performance on D1*

| Group | *n* | Min | Max | Mean | SD | Percentile | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | 25 | 50 | 75 |
| Intervention GR1D1 | 174 | 1.11 | 76.67 | 45.50 | 20.40 | 31.11 | 49.44 | 63.33 |
| Intervention GR2D1 | 174 | 4.44 | 100.00 | 77.94 | 19.74 | 70.00 | 84.44 | 92.22 |

*Note*. All values with the exception of sample size (*N*) are expressed as percentages (%) with a maximum value of 100%.

The box and whisker plot in Figure 4 graphically displays the performance of the intervention group after six months on the programme.

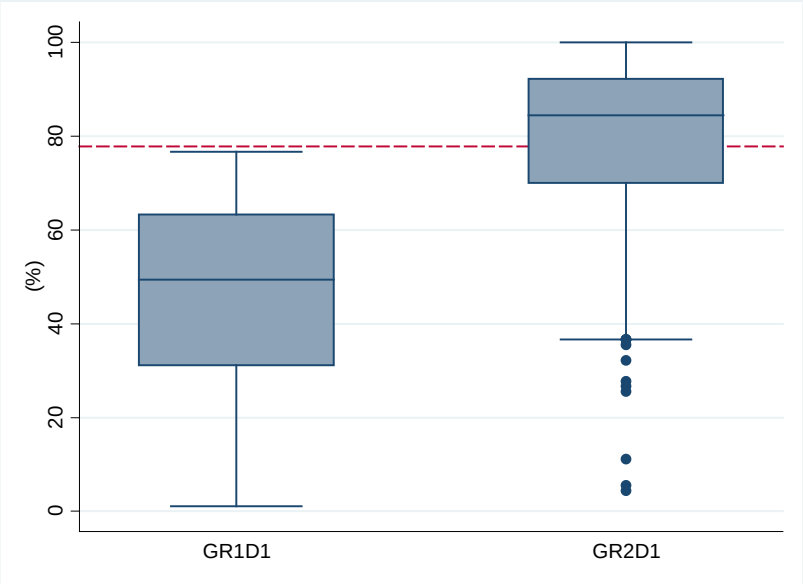


*Figure 4.* Six month box and whisker plots on D1 (*n* = 174)

The dotted red line in Figure 4 indicates Shine's grade appropriate level of 77.77%. 116 out of 174 learners (67%) achieved grade appropriate levels. This is not an official exit point of the programme and therefore even though most of the learners achieved grade appropriate levels, they continued in the second half of the year. Twelve learners did however exit the programme for unknown reasons.

**October 2011 to October 2012 (one year on programme).** Shine uses two tools to measure the literacy ability of the learners in the intervention group after one year on the programme, namely the second diagnostic test (D2) and the reading level test.

***D2 assessment.*** Based on the D2 assessment, out of the total 174 intervention learners who started the programme, 115 (66%) did progress to grade appropriate level after completing one year on the programme. This is slightly less than the 70.00% target set by Shine. As discussed above, this is a conservative calculation as the 12 learners who exited after the first six months were not accounted for in this figure as it was not an official exit point and the reasons for exiting the programme were unknown.

Henceforth I look into the detail of the improvement only for the learners on the programme during the entire time period (i.e., excluding those who exited the programme before the one year assessment). It should be noted at this point that the number of learners in the intervention group is 162, as compared to 174 initially.

Table 5 outlines the summary statistics for learners' scores in the GR1D1, GR2D1 and GR2D2 assessments.

Table 5

*Six Month and One Year Measures of Performance on D1 and D2*

| Group | *n* | Min | Max | Mean | SD | Percentile | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | 25 | 50 | 75 |
| Intervention GR1D1 | 162 | 1.11 | 76.67 | 45.96 | 20.09 | 31.11 | 50.00 | 63.33 |
| Intervention GR2D1 | 162 | 11.11 | 100.00 | 78.37 | 18.10 | 70.00 | 83.89 | 92.22 |
| Intervention GR2D2 | 161 | 16.25 | 100.00 | 79.72 | 19.93 | 71.25 | 87.50 | 93.75 |

*Note*. All values with the exception of sample size (*n*) are expressed as percentages (%) with a maximum value of 100%

The learners' mean score improved from 45.96% to 78.37% over the first six months and their median score improved from 50.00% to 83.89%, an increase of 33.4%1 and 33.89% respectively. However, the learners' mean score only increased from 78.37% to 79.72% over the second six months, and their median score from 83.89% to 87.50% an increase of only 1.35% and 3.61% respectively. This indicates that there was less of an increase in scores in the second

six-month period than in the first. Once again, the range of results in GR1D1, GR2D1 and GR2D2 is very wide.

The box and whisker plot in Figure 5 graphically displays the performance of the intervention learners for all three assessments.

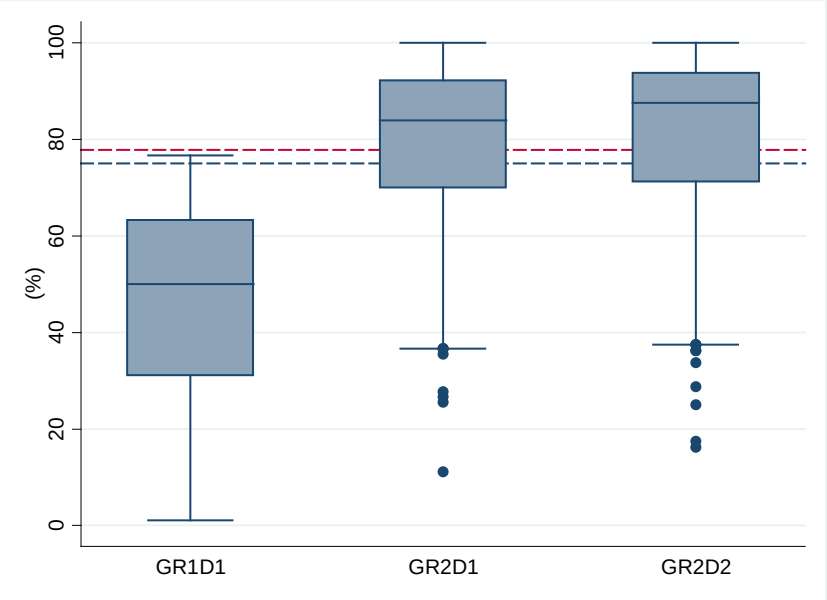


*Figure 5.* Six month and one year box and whisker plots on D1 and D2 (*n* = 162)

As before, the dotted red line in Figure 5 indicates Shines' grade level of 77.77% (a score of 70 out of 90) for GR1D1 and GR2D1. For D2, however, grade level was set at a score of 60 out of 80, or 75.00%. This new Shine grade level is indicated by the dotted blue line. 115 of the 162 (71%) learners who completed one year on the programme achieved Shine grade level on the D2 assessment.

***Reading assessment.*** Based on the first reading assessment, out of the initial 174 intervention learners who started the programme, 118 learners progressed to a grade appropriate reading level of 94 months (7 years and 10 months) after completing one year on the programme, which is greater than the 70.00% target set by Shine. Once again, this does not include the 12 learners who exited the programme at the end of the first six months on the programme.

If one looks at the detail of the 162 learners who completed the full year on the programme, as of October 2012, ten months into Grade 2, the mean score for the intervention group was 99 months. At this stage they are expected to have a reading age of 94 months (7 years and 10 months) and are therefore reading 5 months above the expected reading level.

**October 2011 to May 2013 (18 months on programme).**

***Reading assessment.*** Although Shine did not set an explicit target for this one and a half year assessment point, we assume that it will be 80% which is between the one year target of 70% and the two year target of 90%. In light of this, 134 (77%) of the initial intervention group learners reached the grade appropriate level at the end of one and a half years. This is only slightly below the 80% target and is very close in light of the conservative calculations.

If one looks at the detail of only the learners who were involved in the programme for this period, there are 54 learners in the intervention group. As at May 2013, five months into Grade 3, the learners are expected to have a reading age of 101 months (or 8 years and 5 months). At this point, the mean score for the intervention group was 98.5 months, marginally below grade level. Out of the 54 learners, only 19 learners achieved a reading level of at least 101 months, which is the grade appropriate level.

**October 2011 to October 2013 (two years on programme).** At the end of the two year programme, of the 174 learners who began the Shine programme, 153 (88%) learners reached Shines' grade appropriate level, which is just under the 90% target set by Shine. Once again, this is due to the conservative calculations where learners exit for unknown reasons.

Henceforth, I present the results of only the sample of learners who completed the full two years on the programme. It should be noted that after two years the number of learners in the intervention group was 48, as compared to 174 at the very start. The reduction is because learners who graduated or left the programme previously are excluded from this analysis.

Table 6 outlines the summary statistics for learners' scores in the GR1D1, GR2D1, GR2D2 and GR3D2 assessments.

Table 6

*Six Month, One Year and Two Year Measures of Performance on D1 and D2*

| Group | *n* | Min | Max | Mean | SD | Percentile | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | 25 | 50 | 75 |
| Intervention GR1D1 | 48 | 1.11 | 76.67 | 36.85 | 21.57 | 17.78 | 33.33 | 55.56 |
| Intervention GR2D1 | 48 | 25.56 | 97.78 | 67.48 | 20.17 | 53.33 | 67.78 | 82.78 |
| Intervention GR2D2 | 48 | 16.25 | 98.75 | 63.36 | 21.75 | 45.00 | 68.75 | 78.75 |
| Intervention GR3D2 | 48 | 8.75 | 100.00 | 83.15 | 18.31 | 81.88 | 90.00 | 93.75 |

*Note.* All values with the exception of sample size (*n*) are expressed as percentages (%) with a maximum value of 100%

For this sample, the mean and median score improvement in the first six months was 30.63 and 34.45% respectively. For the second six months of the year the mean and median improvement was -4.14% and 0.97% respectively. From the end of year one to the end of year two, the mean and median improvement was 19.79 and 21.25% respectively. Therefore in total over the two years on the programme the mean and median improvement score was 46.3% and 56.67% respectively.

The box and whisker plot in Figure 6 graphically displays the medians and interquartile ranges for the intervention learners for all four assessments.

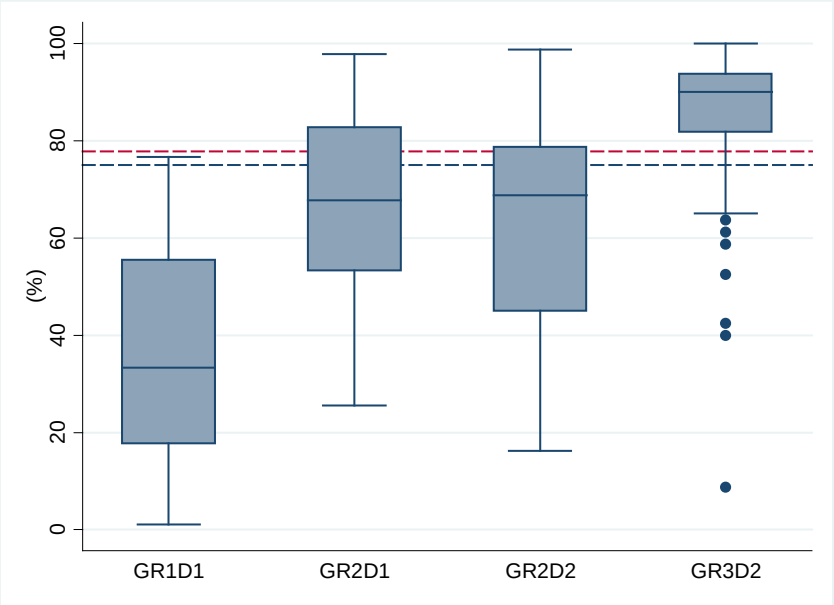


*Figure 6.* Six month, one year and two year box and whisker plots on D1 and D2 (*n* = 48)

The dotted red line indicates Shines' grade level of 77.77% for GR1D1 and GR2D1, while the dotted blue line indicates the grade level of 75.00% for Gr2D2 and GR3D2. Out of the 48 intervention group learners who stayed on the programme for the full two years, 38 achieved scores for their D2 assessment at or above grade level.

In summary of the diagnostic assessments I refer to Table 7 and Figure 7 which illustrates the average changes in the literacy scores of the intervention learners from all the diagnostic assessments in relation to the Shine expected grade level over the full two your period.

Table 7

*Intervention Group Learners' Mean Literacy Scores and Expected Standards*

| | Assessment D1 (Baseline, *n* = 174) | Assessment D1 (6 months, *n* = 174) | Assessment D2 (1 year, *n* = 162) | Assessment D2 (2 years, *n* = 48) |
|---|---|---|---|---|
| Intervention actual | 45.5 | 77.94 | 79.72 | 83.15 |
| Grade appropriate level | 77.77 | 77.77 | 75 | 75 |

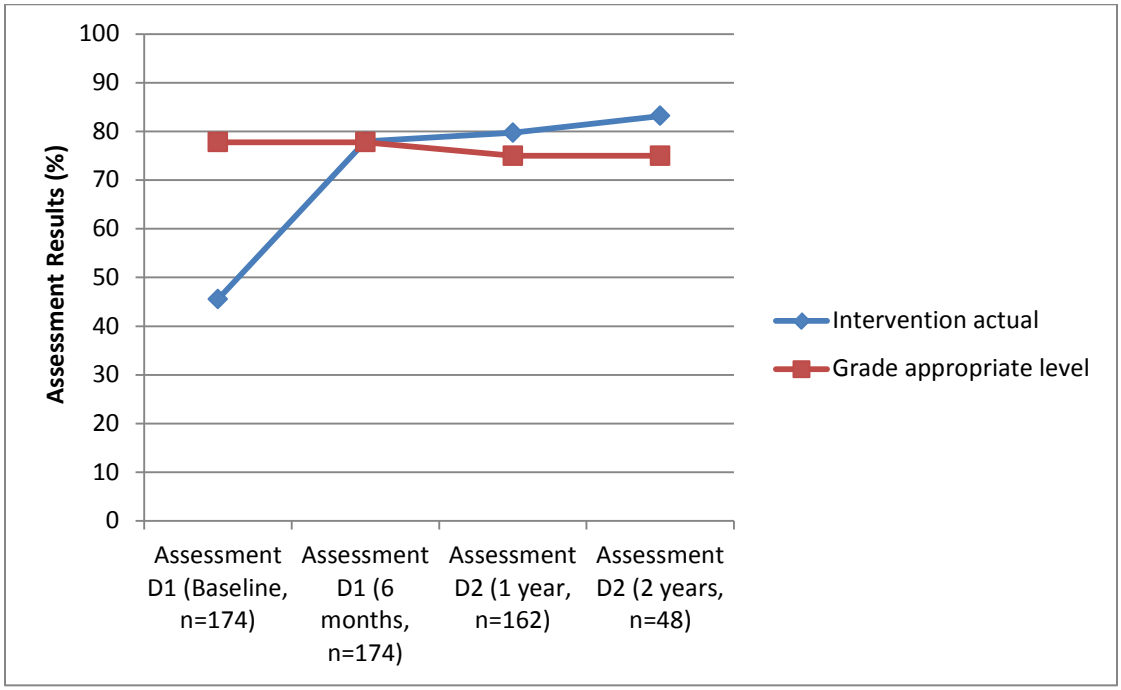


*Figure 7.* Intervention group learners' mean literacy scores and expected grade-appropriate results

It is clear that the greatest improvement occurred over the first six months on the programme. Within the first six month period, most learners reached grade appropriate level. Thereafter, learners continued to improve at a slower rate. At the one year and one and a half year assessment, the learners' average literacy score was greater than grade appropriate level.

In summary of the reading assessments, I refer to Table 8 and Figure 8 which illustrates the intervention learners' mean reading level in comparison to the expected reading level at the point of each assessment.

Table 8

*Intervention Group Learners' Mean Reading Age versus Expected Standards*

| | Mean reading age in months (1 year, *n* = 162) | Mean reading age in months (1.5 years, *n* = 54) |
|---|---|---|
| Intervention actual | 99.31 | 98.53 |
| Grade appropriate level | 94 | 101 |

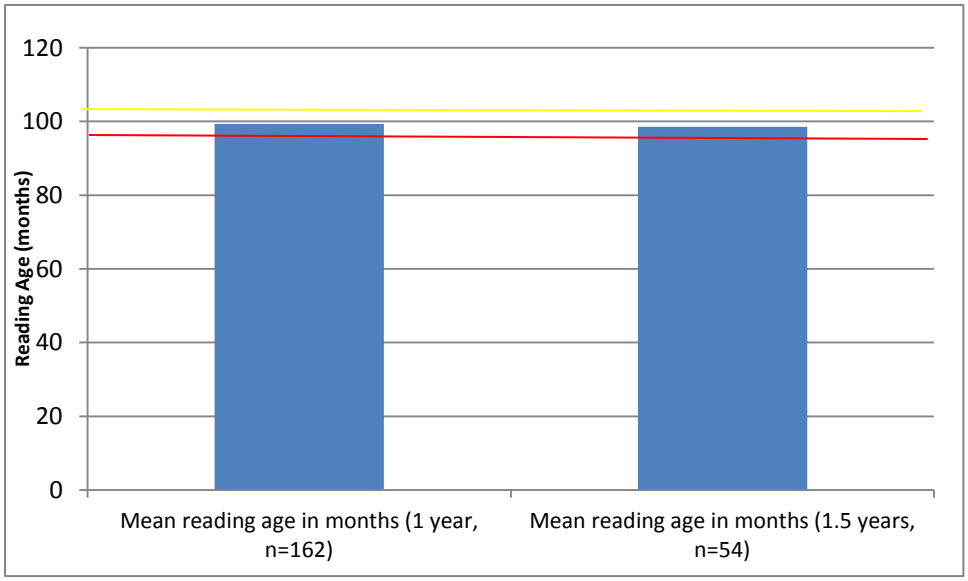


*Figure 8.* Intervention group learners' mean reading scores and expected reading age

In Figure 8, the red line represents the expected reading age at one year (94 months) and the yellow line represents the expected reading age at one and a half years on the programme (101 months). In summary, after one year on the programme, the intervention learners were

reading at 5.31 months above the expected reading age. After one and a half years, on average, learners were reading 2.47 months below the expected reading level.

In summary of the number of graduates throughout the two year programme period, 153 of the 174 learners (88%) who started the intervention below Shines' grade appropriate level and who graduated from the programme, reached Shines' grade appropriate level. This once again is a conservative figure as it only includes learners who graduated from the programme during the official exit points after one or two years.

**Evaluation question 2: Was there any significant difference in the average literacy score of the children who received the programme and those who did not?**

This evaluation question was answered by using a regression discontinuity sampling method. I compared those learners just above the cut-off (77.77%) with those learners just below the cut-off. The data of those two samples were analysed using both parametric and non-parametric models to determine the best fit for the data. Fitting the model correctly is important because it directly influences whether the model can be considered an unbiased estimator of the mean intervention impact around the cut-off point (Jacob et al., 2012). If the model is incorrectly specified then it can introduce a bias. For example, fitting a linear model when the true form is highly non-linear could show a discontinuity when no discontinuity actually exists. For this reason, I will test a variety of functional forms, including both parametric and non-parametric functions. One of the key differences between the two is that for parametric models the evaluator tests pre-determined model forms, whereas for non-parametric models the form is unrestricted.

For the parametric functions, every observation in the sample is used to develop the model. The strength of the parametric method lies in its ability to use observations far away from the cut-off to estimate the average outcome for observations near to the cut-off (Jacob et al., 2012). Hence, this is referred to as a global strategy. Our approach is to use parametric ordinary least squared regression with varying model structures. Moss and Yeaton (2006) explain that although it is preferable to fit the existing data with a linear regression model, since this is simplest, it is important to first rule out the possibility of significant quadratic, cubic, and higher-order polynomial terms as well as potential interactions with these non-linear terms. While I tested all of these models, I have only reported on the model that fits best.

For the non-parametric functions, the analysis is limited to the observations that lie close to the cut-off point. Hence, this is referred to as a local strategy. The range around the cut-off point that is considered for the analysis is referred to as the bandwidth. It is important to note that using a smaller bandwidth typically reduces the sample size, leading to lower statistical power (Jacob et al., 2012). Therefore, for each analysis, the optimal bandwidth will be found using cross-validation, where the model is tested to avoid problems of over fitting.

Question 2 will be answered by considering each of the different time periods which relate to when both the learners in the intervention group and the control group were tested.

As an overview to all the models presented below, I start by considering the two axes in the models. On the x-axis the scores for GR1D1 are indicated, with a vertical dotted line indicating the grade level score of 77.77%. For the 6 month comparison, on the y-axis the scores for GR2D1, with the horizontal dotted line indicating the grade level score of 77.77%. For the one and one and a half year comparison, the y-axis indicates the learner's reading age in months. The horizontal lines indicate 94 months for the one year comparison and 101 months for the one and a half year comparison. For all the models, learners to the left of the vertical dotted line are those in the intervention group (pink line), while those to the right are in the control group (blue line).

**October 2011 to May 2012 (six months on programme).** The best fitting parametric model was a linear model with interaction terms, illustrated in Figure 9.

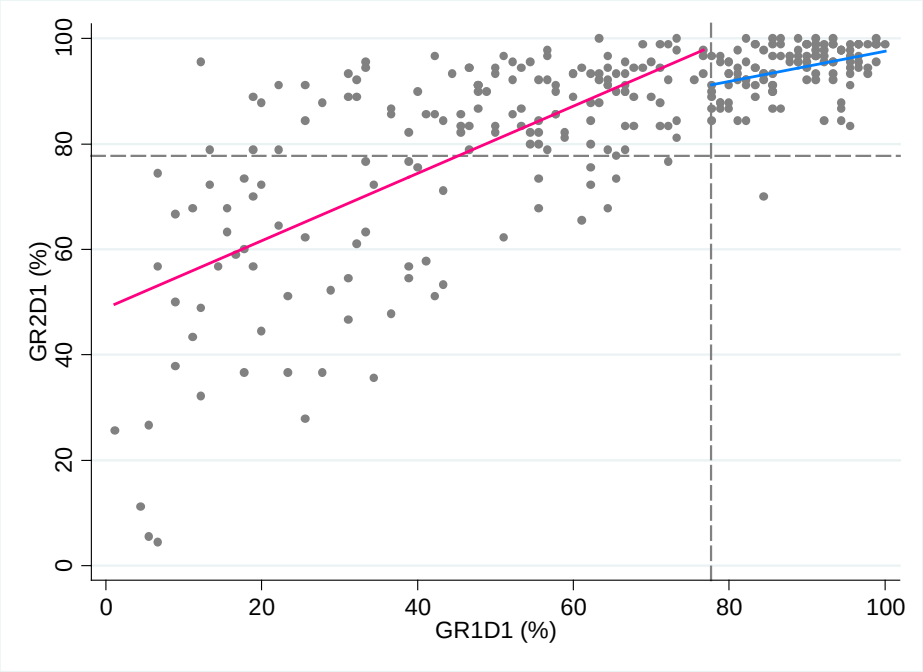


*Figure 9.* Six month parametric model on D1 ($n_{intervention}$ = 174; $n_{control}$ = 121)

The slopes of the lines for the two groups are markedly different. This suggests that there is some difference between the increases in learners' scores from GR1D1 to GR2D1 based on the group to which they belong.

If I look at this in more detail, I see that blue line shows a weak positive relationship. For every 1.00% increase in GR1D1, GR2D1 increases by 0.29% on average. The pink line shows a stronger positive relationship, with a 0.64% increase in GR2D1 for every 1.00% increase in GR1D1. I also notice that pink and blue lines intersect the vertical dotted line at different points, which means that the outcomes are different. The difference between the two points is 7.36%. The fact that the pink line intersects higher than the blue line indicates that the intervention group learners do indeed have a greater improvement in literacy scores from GR1D1 to GR2D1 than the control group learners. This discontinuity is statistically significant at the 5% level which means that the change reflected in the learners' results is as a direct result of the intervention.

The non-parametric model is illustrated in Figure 10.

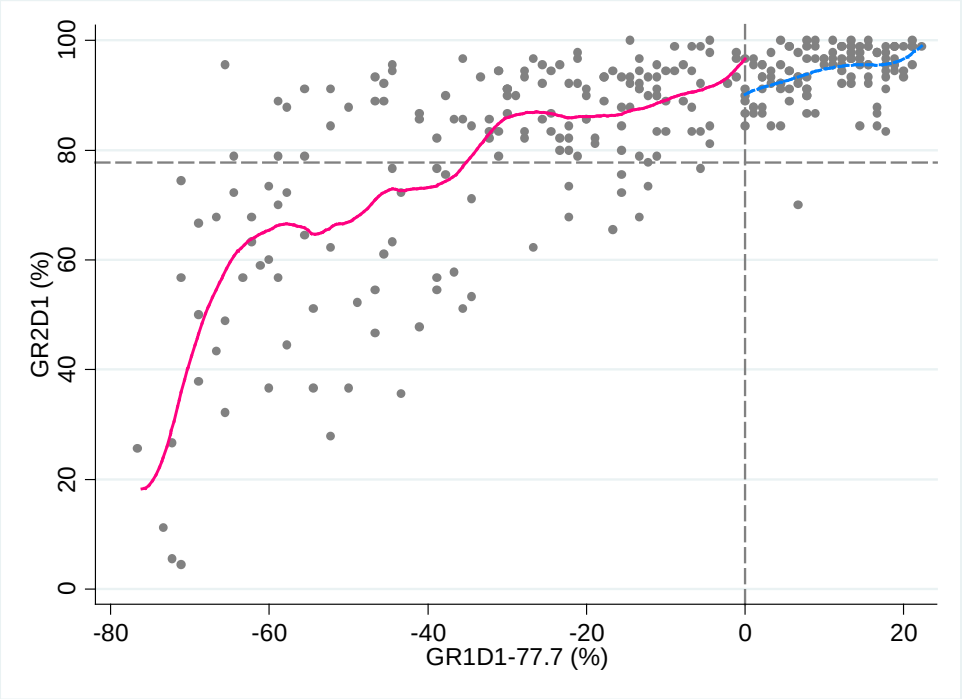


*Figure 10.* Six month non-parametric model on D1 ($n_{\text{intervention}}$ = 174; $n_{\text{control}}$ = 121)

As with the parametric model, there is a positive discontinuity at the vertical dotted line that represents a difference in the outcome level. This discontinuity of 6.47% is statistically

significant at the 5% level. This means that the change reflected in the learners' results is as a direct result of the intervention.

Both the parametric and non-parametric models support the conclusion that the intervention group learners did achieve a statistically significantly greater improvement than control group learners in their literacy score.

The details of the parametric and non-parametric models are outlined in Table 9.

Table 9.

*Output From Six Month D1 Regression Discontinuity Models*

| Variable | Parametric Model | | | | Non-Parametric Model | | | |
|---|---|---|---|---|---|---|---|---|
| | B | SE(B) | *t* | *p* | B | SE(B) | Z | *p* |
| Group Effect | 7.36 | 2.70 | 2.73 | .007 | | | | |
| GR1DI Effect (Control) | 0.29 | 0.17 | 1.70 | .089 | | | | |
| Change in GR1D1 Effect | 0.35 | 0.7 | 2.01 | .046 | | | | |
| Constant | 91.19 | 2.11 | 43.16 | .000 | | | | |
| Lwald | | | | | -6.47 | 1.92 | -3.21 | .001 |

*Note*. A positive coefficient in the parametric model is consistent with a negative coefficient in the non-parametric model. This indicates that the intervention group has a higher value at the cut-off.

**October 2011 to October 2012 (one year on programme).** The best fitting parametric model was a linear model with interaction terms, illustrated in Figure 11 .

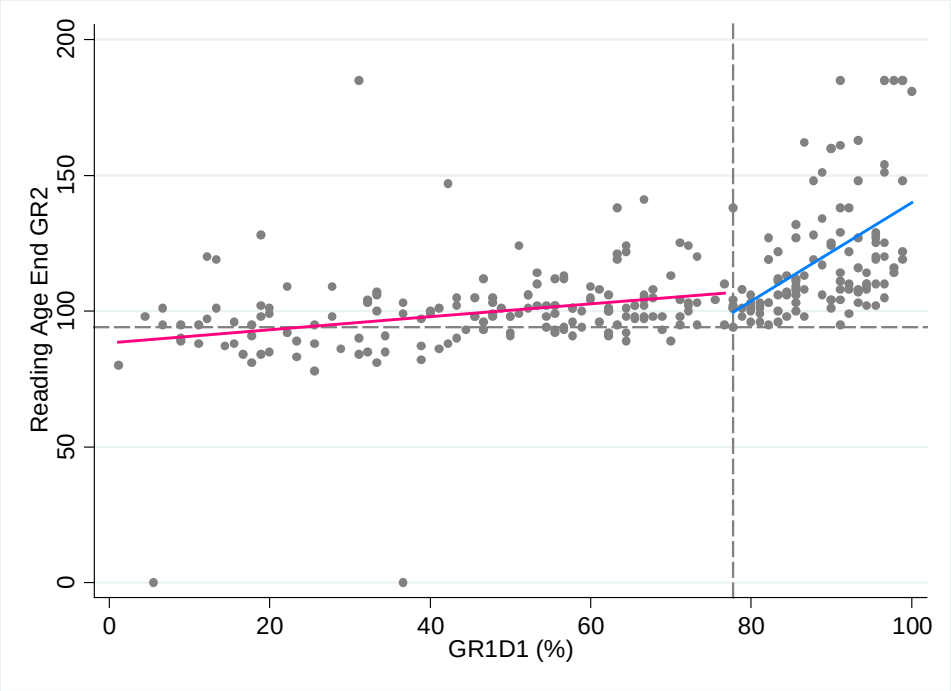


*Figure 11*. One year parametric model on D1 and GR2 Reading ($n_{intervention}$ = 162; $n_{control}$ = 117)

If we look at this in more detail, we see that the blue line shows a strong positive relationship. For every 1.00% increase in GR1D1, reading age increases by 2.01 months. The pink line shows a weaker positive relationship, with a 0.26 month increase in reading age for every 1.00% increase in GR1D1.

We also notice that the pink and blue lines intersect the vertical dotted line that represents grade level at different points. The difference between the two points is 7.25 months. However, this discontinuity is not statistically significant at the 5% level. This means that we cannot conclude that the change reflected in the learners' results is as a direct result of the intervention at a 5% significance level. However, the discontinuity is significant at a 10% level.

The non-parametric model is presented in Figure 12.

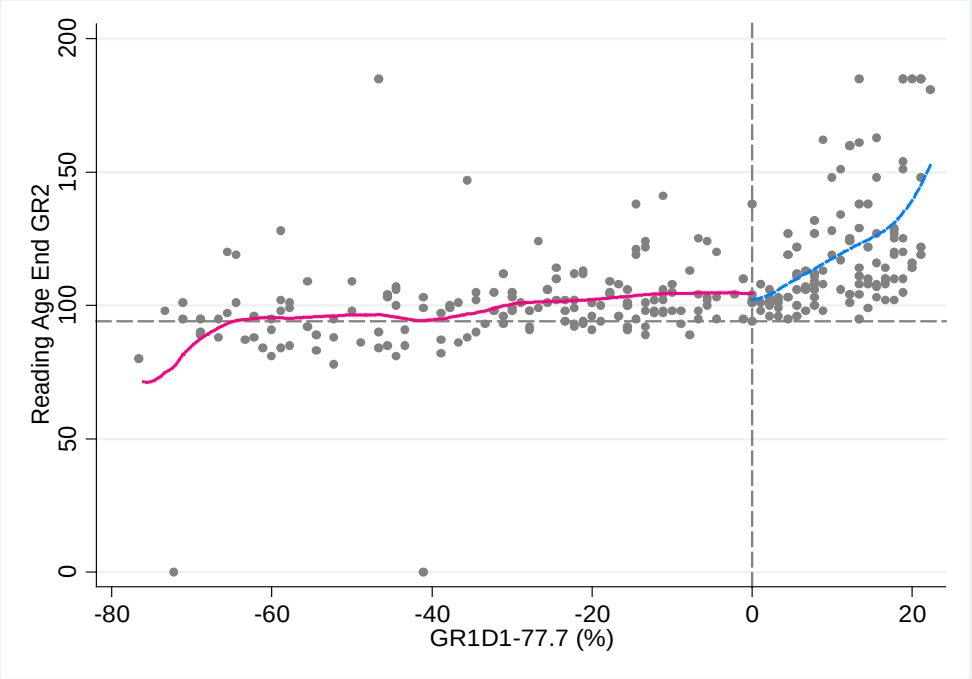


*Figure 12.* One year non-parametric model on D1 and GR2 Reading ($n_{intervention}$ = 162; $n_{control}$ = 117)

As with the parametric model, there is a positive discontinuity at the vertical dotted line that represents a difference in the outcome level. This discontinuity of 2.87 months is not statistically significant at the 5% level. This means that we cannot conclude that the change reflected in the learners' results is as a direct result of the intervention. The discontinuity on the non-parametric model is also not significant at a 10% level.

Both the parametric and non-parametric models support the conclusion that the intervention group of learners did not achieve a statistically significantly greater improvement than control group learners in their reading level scores at a 5% significance level. Even though the non-parametric model does not show a significant difference at the 10% level, the parametric model does. This allows one to draw the conclusion that the discontinuity is significant at the 10% level. This means that at the 10% significance level, the intervention group of learners did achieve a statistically significantly greater improvement than control group learners in their reading level scores, which can be attributed to the programme.

The details of the parametric and non-parametric models are outlined in Table 10.

Table 10

*Output from One Year D1 and GR2 Reading Regression Discontinuity Models*

| | Parametric Model | | | | Non-Parametric Model | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | B | SE(B) | t | p | B | SE(B) | Z | p |
| Group Effect | 7.25 | 4.27 | 1.70 | .091 | | | | |
| GR1DI Effect (Control) | 2.01 | 0.29 | 6.89 | .000 | | | | |
| Change in GR1D1 Effect | -1.75 | 0.30 | -5.77 | .000 | | | | |
| Constant | 99.70 | 3.32 | 30.06 | .000 | | | | |
| Lwald | | | | | -2.87 | 5.21 | -0.55 | .582 |

Note: A positive coefficient in the parametric model is consistent with a negative coefficient in the non-parametric model. This indicates that the intervention group has a higher value at the cut-off.

**October 2012 to May 2013 (18 months on programme).** The best fitting parametric model illustrated in Figure 13 was a linear model with interaction terms.

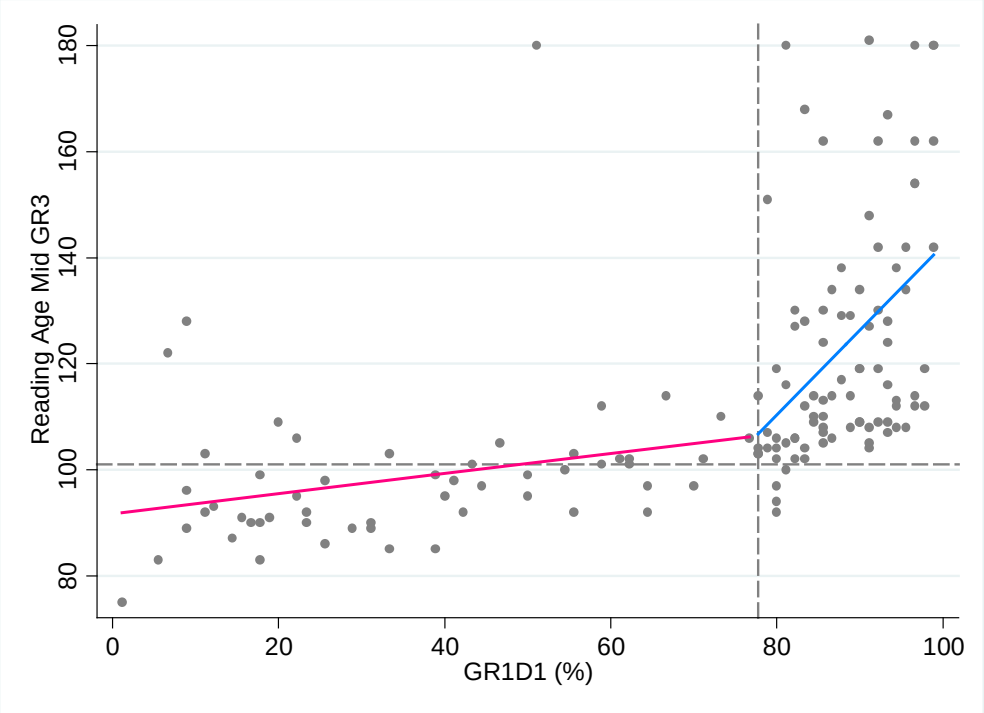


*Figure 13.* One and a half year parametric model on D1 and GR3 Reading ($n_{intervention}$ = 54; $n_{control}$ = 89)

We see that the blue line shows a strong positive relationship. For every 1.00% increase in GR1D1, reading age increases by 1.78 months. The pink line shows a weaker positive relationship, with a 0.21 month increase in reading age for every 1.00% increase in GR1D1.

We also notice that the pink and blue lines intersect the vertical dotted line, showing the difference in outcome levels. The difference between the two points is 0.31 months, and this discontinuity is not statistically significant at the 5% level. This means that we cannot conclude that the change reflected in the learners' results is as a direct result of the intervention. Furthermore, the discontinuity is not significant at a 10% significance level.

The non-parametric model is illustrated in Figure 14.

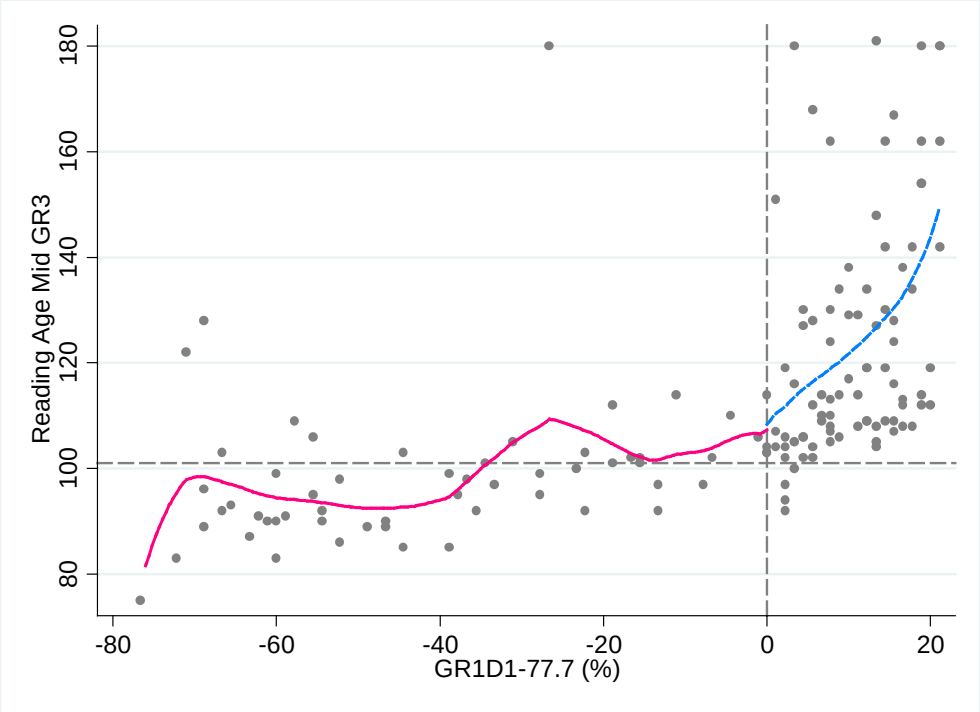


*Figure 14.* One and a half year non-parametric model on D1 and GR3 Reading ($n_{intervention}$ = 54; $n_{control}$ = 89)

As with the parametric model, there is a very slight discontinuity at the vertical dotted line that represents grade level. This discontinuity of 1.65 months is not statistically significant at the 5% level. This means that we cannot conclude that the change reflected in the learners' results is as a direct result of the intervention.

Both the parametric and non-parametric models support the conclusion that the intervention group learners did not achieve a statistically significantly greater improvement than control group learners in their reading level scores.

The details of the parametric and non-parametric models are outlined in Table 11.

Table 11

*Output from Six Month D1 and GR3 Reading Regression Discontinuity Models*

| Variable | Parametric Model | | | | Non-Parametric Model | | | |
|---|---|---|---|---|---|---|---|---|
| | B | SE(B) | *t* | *p* | B | SE(B) | Z | *p* |
| Group Effect | -.31 | 6.73 | -0.05 | .963 | | | | |
| GR1DI Effect (Control) | 1.78 | 0.35 | 5.10 | .000 | | | | |
| Change in GR1D1 Effect | -1.57 | 0.37 | -4.20 | .000 | | | | |
| Constant | 106.72 | 3.79 | 28.19 | .000 | | | | |
| lwald | | | | | 1.65 | 4.99 | 0.33 | .740 |

*Note*. A positive coefficient in the parametric model is consistent with a negative coefficient in the non-parametric model. This indicates that the intervention group has a higher value at the cut-off.

In summary, the results demonstrated that after six months the improvement in the learner results was as a direct result of the literacy hour programme at a 5% significance level. After one year, although the difference in the results of the intervention and control group did not show significance at a 5% level, it did at a 10% level. At the one and a half year reading assessment, I could not conclude that the improvement was a result of the programme at either the 5% or 10% significance level. This result resonates with the literature and was therefore not unexpected.

## DISCUSSION

The results will be discussed in reference to each of the evaluation questions and the relevant assessment time periods.

**Evaluation Question 1: Did the literacy scores for the children on the programme improve to grade appropriate levels and did Shine reach their targets?**

The progress in the literacy results of the intervention learners was positive. Shine reached or just missed (with conservative calculations) its targets in terms of the number of the original intervention group learners who reached grade level over the two year time period on the programme.

The average results for the diagnostic assessments were above the expected grade level at each time period. For the reading assessment the average reading level of the learners was above grade level at the one year point and just below at the one and a half years point. By examining the detail of these results we can better understand them.

After only six months on the programme the majority of the intervention learners (67%) progressed to Shines' grade appropriate level on the D1 assessment. The average improvement for the learners who completed six months on the programme was 32%. This result exceeded Shine's target, which only set an expected average improvement of 22%. This illustrates a very steep learning curve and a high level of success for the Shine learners in this six month period.

After one year on the programme, learners were tested using the D2 assessment and the first reading assessment. According to the D2 assessment the cumulative percentage of learners reaching grade level (out of the original 174) was 66%. According to the reading assessment the cumulative percentage of learners reaching grade level (out of the original 174) was 68%. Both these results are just below the target of 70% set by Shine. Objectively, this result is a success. Even though it is just below the target, this is only because of the conservative calculations, as explained in the results chapter.

With regards to the 162 learners who completed the full year on the programme, the overall mean results increased to the expected grade level or above for both the D2 assessment and the reading assessment. Even though the increase in learner results for the D2 assessment was less in the second six months of the year, this is understandable as most had already reached

grade appropriate level and the D2 assessment was set at a higher standard than the D1 assessment. For the reading assessment, the mean reading score was 5 months above the expected reading level. This means that these learners were successful in reaching grade appropriate level for the literacy skills they should have accomplished by the end of Grade 2 and they had also transferred those skills into reading skills.

After one year, the majority of the learners had reached grade level and graduated from the programme. The learners who continued on the programme into the second year were the weaker subset of the original intervention group learners.

After one and a half years the average reading level according to the reading assessment fell below the grade appropriate level. Only 19 of the 54 learners (35%) who attended the programme for one and a half years achieved grade appropriate levels. This result is not unexpected considering that these were the weaker learners in the sample.

This result resonates with the literature. Elbaum et al. (2000) note that for learners with severe difficulties in reading, the average effect size of a literacy intervention would likely be insufficient to raise their performance to grade appropriate level or to eliminate the need for sustained support beyond the time frame of the intervention.

However, after two years on the programme 38 out of the 48 weaker learners achieved grade appropriate level on the D2 assessment.

It is also interesting that many more learners failed to reach grade level in the reading assessment after one and a half years than in the literacy assessment after two years. Had the reading assessment taken place after two years, we could reasonably expect the results to have been somewhat better, although they would likely have still been worse than for the literacy skill assessment (D2). This is consistent with the conclusions of Cadieux and Boudreault (2005) who note in their study that Grade 1 learners, when receiving a one-to-one intervention, made significant gains in phonetic awareness, but not in reading. They add that although these learners may not reach grade appropriate levels in all areas, they likely benefit enough from the intervention to keep up with classroom instruction and to avoid academic failure. Consequently, while these learners failed to reach grade level in reading, it might be the case that their improvements relative to their very low starting points can still be considered satisfactory.

In summary, the majority of the intervention learners who started the programme below grade level reached grade appropriate levels in both literacy skills, as tested by the diagnostic tests, and in reading, as tested by the reading test. This is a massive accomplishment for Shine.

The small sample size per school in this evaluation limited my ability to perform statistically credible inter-school comparisons. However, the observations below indicate that future studies in this particular area could provide valuable insights.

The inter-schools results reflected some of the trends in the literature on South African literacy results. As explained by Hungi (2011) one of the main predictors of learner failure in South Africa is socio-economic status. The author points to the results from the 2000 and 2007 SACMEQ reports indicating that learners from wealthy schools significantly outperformed their counterparts from poor schools. School-specific predictors include school resources (Hungi, 2011). In the context of this evaluation, the proxies for socio-economic status are school quintile, and fee per learner per year. In general we would expect a school with a higher socio-economic status to have a higher quintile, and a school with a lower socio-economic status to have a lower quintile and lower fee per learner per year. School 2 and School 6 are the two schools in the sample with Shine Chapters rather than Shine Centres. They are also the only two quintile 4 schools. The other schools are quintile 5. From the results we see that after one year School 6 and School 2 had the lowest graduation proportion of all the schools, while after two years School 2 had the lowest graduation proportion. Additionally we saw that the number of learners that reached grade appropriate level at the numerous assessment points varied significantly between schools. For example, 88% of the 51 learners that started the programme at School 3 graduated after one year. By comparison, only 65% of the 43 learners that started the programme at School 4 graduated after the first year. If we consider that the average fee per learner is roughly four times higher at School 3 than at School 4 then this result is perhaps unsurprising.

Considering that the Shine literacy assessment is performed in English, it is plausible to assume that the language of tuition could also potentially explain some of the variation between schools. School 2 is the only school in the sample that is bilingual, meaning that teaching takes place partially in English and partially in Afrikaans. This school had the lowest percentage of learners graduating after the full two years on the programme.

**Evaluation Question 2: Was there any significant difference in the average literacy score of the children who received the programme and those who did not?**

In answering the second question, I analysed the difference between the improvement of the intervention learners who were just below the cut-off point and the control group learners who were just above the cut-off point. This allowed me to determine if any improvement by the intervention learners, which we have seen in the results of the first question, was a direct result of the Shine programme.

After the first six months on the programme there was a significantly greater improvement for the learners in the intervention group than for those in the control group. This demonstrates that not only did the improvement of the intervention group of learners improve by 32% but also that this improvement was a direct result of the programme. This means that for learners who are at a similar literacy level at the end of Grade 1, one would expect only those who attended the Shine programme for six months to achieve these results, all other things being equal.

However, after one year and one and a half years on the programme the results reflected no statistically significant difference between the intervention and control groups at a 5% significance level. This was despite there being a clear positive difference of 7.25 months, a difference of more than half a year of reading improvement. It is surprising that despite this increase we are unable to conclude at the 5% level that the Literacy Hour programme led to these improvements in the literacy scores. There are several potential explanations for this. The most obvious is that the limited sample size meant that there was insufficient power to find a significant result at this level. The improvement level of the intervention group was however significantly greater than the control group learners at the 10% significance level. It is therefore important to ask what level of statistical significance is relevant in the literacy sector and for Shine in particular.

For the reading level assessment after one and a half years on the programme, the intervention learners showed no statistically significant difference in their results when compared to the control group at either the 5% or 10% significance level. As such, the difference in results could not be attributed to the intervention. Although this seems more reasonable, as the learners

on the programme at this point are the very weakest of the intervention learners, the small sample size could still have had an impact.

**Limitations**

There are several limitations that qualify the conclusions drawn above. The primary limitation relates to the sample size. In general, the larger the sample size, the greater the likelihood that conclusions drawn from the sample will be considered statistically credible. As outlined in the discussion chapter, we only considered learners from the 2011 to 2013 cohort, from Western Cape schools, and for whom classifications into the intervention and control groups were accurate throughout the two years. There were also a handful of learners who started the Literacy Hour programme but dropped out for unknown reasons. While the number of learners in the intervention group was initially 174, there were only 48 learners who participated in the final assessment after two years. Although the sample size did decrease dramatically, it demonstrates the positive results of the programme as learners were reaching grade level and graduating the programme. However, the smaller sample size does affect the power of the evaluation to find statistically significant results. In addition, it is not possible to draw statistically credible conclusions about differences between subsets within the data, for example between schools or between Centres and Chapters.

Another limitation relates to the measurement tools. There are several parts to this. The first issue is the credibility and reliability of the assessment measures. The assessments have not been externally verified against the Department assessment standards. This specifically relates to how these assessments compare to the standard expected by the Department of Basic Education. One needs to question the appropriateness of Shine's cut-off point, (i.e. the grade appropriate level). Is Shine's 77.77% cut-off grade level equal to the Departments 50% pass level? It is important to know that if learners achieve grade appropriate level on the Shine assessment, it means that they should achieve grade appropriate level in the classroom. This verification needs to be considered in future assessments.

The third limitation is that the assessments have been through several changes over the years. This complicates inter-cohort comparisons which are often very valuable from a monitoring and evaluation perspective in determining whether an intervention makes a sustained impact. This also limits the number of learners that can be included in the sample.

In addition, the application of the measurement tools is inconsistent. Specifically, the control group learners are not tested at all using the D2 assessment. In addition, the final reading comparison is at one and a half years, which does not allow for the necessary comparison at two years. This affects one's ability to compare learners at all the assessment stages.

Lastly, the details for the various sub-components of the assessments are not recorded. These sub-components could include reading, writing, comprehension and phonics. The fact that we only see an overall mark gives the evaluation a black-box effect. From a programme theory perspective, this makes it difficult to determine which components of the programme work, or to understand why the programme does or does not work.

These limitations of the evaluation do not undermine the results presented in this dissertation but rather highlight some possible improvements for future evaluations.

**Recommendations**

On the basis of the discussion and limitations discussed above, I am able to outline a number of recommendations.

With regards to the measurement tools, it is recommended that each of these assessments should be validated against DBE standards, and should be completed by all learners, regardless of whether they were initially in the intervention or control groups. The assessments should be kept consistent for as long as they remain relevant, with changes only being made where absolutely necessary. As a direct consequence of the above, Shine will be able to increase the number of learners in the sample for future evaluations.

I recommend that Shine should monitor the allocation of appropriate learners to the programme, as well as the data capturing more carefully. Extra training of school coordinators in data capturing is an option to help avoid the large number of missing data and incorrect categorisations.

I also recommend that volunteers at schools who have extra capacity to take on learners who do not need the programme according to the Shine Assessment, should move to less resourced schools where learners are in need but cannot receive the programme due to limited

capacity. This will help to increase the sample size as learners will be selected and classified appropriately.

It is also recommended that the details relating to each sub-component of the assessments, along with details of the number of sessions attended by each learner, be recorded comprehensively and consistently. This would allow an evaluator to more thoroughly assess the impact of the intervention. In addition to this, I recommend that future evaluations take into account the dosage of the programme that each learner obtains and the effect thereof.

Finally, it is recommended that Shine monitors and evaluates whether the learners who have graduated from the programme remain at grade appropriate level in the short to medium term after they complete the programme or if there is a drop-off in results.

## CONCLUSION

The South African literacy results in the most recent international studies have been a great cause of concern. As discussed in the introduction chapter, in the 2006 PIRLS South Africa scored the lowest out of the 45 participating countries (Spaull, 2013). In the easier prePIRLS, where learners were tested in their home language, 71% of the Grade 4 learners only reached the Low international benchmark (Howie et al., 2011). For the SACMEQ III, the proportion of South African Grade 6 learners that were deemed functionally illiterate was 27% in comparison to the average of 18% for all participating countries (Spaull, 2012). South Africa ranked tenth out of the 15 participating countries in mean reading score (Spaull, 2011).

Although Shine's assessments cannot be directly compared to these assessments, one can compare the overall pictures. The learners who started in the Shine programme were classified below grade level, fitting well into the worrying picture of South African literacy from the international reports. However, after only six months on the Shine programme most learners demonstrated results at or above grade level and this was as a direct result of the Literacy Hour programme. This represents a remarkable turnaround in the learner's achievement level.

Once again, after one year on the programme, most of the learners kept up their achievement in reaching grade level. Although the difference in the results of the intervention and control group did not show significance at a 5% level, it did at a 10% level. The most appropriate way forward is to increase the sample in future evaluations and to retest the 5% significance level.

Although not all the weakest learners in the sample reached grade appropriate level in the reading test after one and a half years, the majority did reach grade appropriate level on the skills test after two years on the programme. At the one and a half year reading assessment, I could not conclude that the improvement was a result of the programme at either the 5% or 10% significance level. However, this resonates with the literature and was therefore not unexpected.

In conclusion, it is clear that Shine's Literacy Hour programme has a very successful foundation from which to work. While there are opportunities to improve certain elements, as outlined in the recommendations, there is no doubting the fact that learners are measurably

benefiting from the programme. This is a credit not only to Shine but also to the schools, teachers and volunteers who play a part in enabling the learners to flourish.

In the context of South Africa's very serious literacy challenge, Shine is giving struggling learners an opportunity to become the positive exceptions in this story. In so doing, it is actively playing a part in contributing to a nation of readers.

## REFERENCES


Begeny, J. C., Ross, S. G., Greene, D. J., Mitchell, R. C., & Whitehouse, M. H. (2012). Effects of the Helping Early Literacy with Practice Strategies (HELPS) reading fluency program with Latino English language learners: A Preliminary Evaluation. *Journal of Behavioral Education*, *21*(2), 134–149. doi:10.1007/s10864-011-9144-7

Bodrova, E., & Leong, D. (1998). Scaffolding emergent writing in the zone of proximal development. *Literacy Teaching and Learning*, *3*(2), 1–18. Retrieved from http://www.earlyliteracyinfo.org/documents/pdf/doc_122.pdf

Cadieux, A., & Boudreault, P. (2005). The effects of parent-child paired reading program on reading abilities, phonological awareness and self-concept of at-risk pupils. *Reading Improvement*, (410), 224–237. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:THE+EFFECTS+OF+A+PARENT-CHILD+PAIRED+READING+PROGRAM+ON+READING+ABILITIES+,+PHONOLOGICAL+AWARENESS+AND+SELF-CONCEPT+OF+AT-RISK+PUPILS+^#0

Department of Basic Education. (2013). *Report on the annual national assessment of 2013*. Retrieved from www.education.gov.za/LinkClick.aspx?fileticket=Aiw7HW8ccic=

Ehri, L. C., Dreyer, L. G., Flugman, B., & Gross, a. (2007). Reading Rescue: An effective tutoring intervention model for language-minority students who are struggling readers in first grade. *American Educational Research Journal*, *44*(2), 414–448. doi:10.3102/0002831207302175

Elbaum, B., Vaughn, S., Tejero Hughes, M., & Watson Moody, S. (2000). How effective are one-to-one tutoring programs in reading for elementary students at risk for reading failure? A meta-analysis of the intervention research. *Journal of Educational Psychology*, *92*(4), 605–619. doi:10.1037//0022-0663.92.4.605

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology*, *104*(4), 879–896. doi:10.1037/a0029185

Howie, S., van Staden, S., Tshele, M., Dowse, C., & Zimmerman, L. (2011). Progress in international reading literacy study 2011: South African children's reading literacy achievement summary Report. Centre for Evaluation and Assessment. University of Pretoria. Retrieved from http://web.up.ac.za/sitefiles/File/publications/2013/PIRLS_2011_Report_12_Dec.PDF

Hungi, N. (2011). Accounting for Variations in the Quality of Primary School Education. *Working Paper*, (September). Retrieved from http://www.sacmeq.org/sites/default/files/sacmeq/reports/sacmeq-iii/working-papers/07_multivariate_final.pdf

Jacob, R., Somers, M., Bloom, H., Stange, K., Smith, J., Weiss, M., & House, E. (2012). A practical guide to regression discontinuity. Retrieved from http://www.mdrc.org/sites/default/files/regression_discontinuity_full.pdf

Khosa, G. (n.d.). A long road ahead, but the flame has been lit. Retrieved from http://nect.org.za/news/a-long-road-ahead-but-the-flame-has-been-lit

Lake, C., Slavin, R., Susan, D., & Nancy, M. (2009). What works for struggling readers: A best synthesis. This review is adapted from Slavin, R.E., Lake, C., Davis, S., & Madden, N. (2009), Effective Programs for Struggling Readers: A Best Evidence Synthesis, Baltimore MD: Johns Hopkins University, Center for Research and Reform in Education (CRRE). The full report is available at www.bestevidence.org.uk.Retrieved from www.bestevidence.org.uk

Lesik, S. A. (2006). Applying the regression-discontinuity design to infer causality with non-random assignment. *The Review of Higher Education*, *30*(1), 1–19. doi:10.1353/rhe.2006.0055

Moss, B. G., & Yeaton, W. H. (2006). Shaping policies related to developmental education: An evaluation using the regression-discontinuity design. *Educational Evaluation and Policy Analysis*, *28*(3), 215–229. doi:10.3102/01623737028003215

Mullis, I.VS., Martin, M.O., Foy, P. & Drucker, K.T (2011). PIRLS 2011 International results in reading. TIMMS AND PIRLS International Study Center, Lynch School of Education Boston College. Retrieved from http://eric.ed.gov/?id=ED544362

Perraillon, M. C. (2013). Regression discontinuity design. (PowerPoint Slides) University of Chicago. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/0470013192.bsa550/full

Phillips, L. M., Norris, S. P., & Anderson, J. (2008). Unlocking the door: Is parents' reading to children the key to early literacy development? *Canadian Psychology/Psychologie Canadienne*, *49*(2), 82–88. doi:10.1037/0708-5591.49.2.82

Piasta, S. B., Justice, L. M., McGinty, A. S., & Kaderavek, J. N. (2012). Increasing young children's contact with print during shared reading: longitudinal effects on literacy achievement. *Child Development*, *83*(3), 810–20. doi:10.1111/j.1467-8624.2012.01754.x

Rasinski, T. V, Fredericks, A. D., & Rasinski, V. (1991). Working with parents: The Akron Paired Reading Project. *The Reading Teacher*, *44*(7), 514–515. Retrieved from http://www.jstor.org/stable/20200718 .

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). Evaluation: A systematic approach (7th ed.). Thousand Oaks, CA: Sage.

Schochet, P., Cook, T., Deke, J., Imbens, G., Lockwood, J.R., Porter, J., Smith, J. (2010). Standards for regression discontinuity designs. Retrieved from What works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_rd.pdf.Spaull, N. (2012). *South Africa at a glance. SACMEQ at s glance series. Research on Socio-economic Policy (RESEP).* Retrieved from http:resp.sun.ac.za/index.php/projects/

Shadish, W.R., Cook, T.D. and Campbell, D.T. (2002). *Experimental and quasi-experimental design for generalized causal inference*. Houghton Mifflin Company (U.S.A).

Spaull, N. (2011) – A preliminary analysis of SACMEQ III South Africa. Stellenbosch Economic Working Papers: 11/11

Spaull, N. (2012). South Africa at a glance. SACMEQ at a glance series. Research on Socioeconomic Policy (RESEP). Available: http://resep.sun.ac.za/index.php/projects/

Spaull, N. (2013). South Africa's education crisis: The Quality of Education in South Africa 1994–2011, *27*(October). Retrieved from http://www.section27.org.za/wp-content/uploads/2013/10/Spaull-2013-CDE-report-South-Africas-Education-Crisis.pdf

Spaull, N. (2014). Accountability in South African Education. Ch4 in "Transformation Audit 2013: Confronting Exclusion" p. 51-65 Institute for Justice and Reconciliation. Cape Town. Retrieved from http://ijr.org.za/publications/pdfs/TA%202013%20text%20and%20cover%20web.pdf

Stuart, M. (1999). Getting ready for reading: Early phoneme awareness and phonics teaching improves reading and spelling in inner-city second language learners. *British Journal of Educational Psychology*, *69*(4), 587–605. doi:10.1348/000709999157914

Topping, K. J. (1995). *Paired reading, spelling, and writing: The handbook for teachers and parents*. Cassell (London and New York, NY).

Wasik, B. A., & Slavin, R. (1993). Preventing early reading failure with one-to-one tutoring: A review of five programs. *Reading Research Quaterly*, *28*(2), 178–200. International Reading Association. Retrieved from http://www.jstor.org/stable/747888 .

Wasik, B. A. (1998). Using volunteers as reading tutors: guidelines for successful practices. *The Reading Teacher*, *51*(7), 562–570. International Reading Association. Retrieved from http://www.jstor.org/stable/20201962 .

Weissenberg, M.G., Torres, K., Nadler-Nir, E., O'Carroll, S., James, S., Heurmann, M. & Van Durren, L. (2014) The Shine Centre Training Manual for Shine Learning Partners. The Shine Centre (Cape Town)

Zucker, T. A, Cabell, S. Q., Justice, L. M., Pentimonti, J. M., & Kaderavek, J. N. (2013). The role of frequent, interactive prekindergarten shared reading in the longitudinal development of language and literacy skills. *Developmental Psychology*, *49*(8), 1425–39. doi:10.1037/a0030347