An Outcome Evaluation of the LifeMatters Foundation's Numeracy Programme

Nick Hardwick

(HRDNIC011)

A dissertation submitted in partial fulfilment of the requirements for the award of the

Degree of Master of Philosophy (Programme Evaluation)

Faculty of Commerce

University of Cape Town

2017

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: Date: 27/01/2017

Acknowledgments

I would like to thank the following people, because without them this dissertation may never have been submitted:

My supervisor, Joha Louw-Potgieter, for her constant support and timely feedback, and for her patience with me to the end.

Louise Ansley, for her communication and help during the dissertation process.

For all the staff I dealt with at the LifeMatters Foundation, you made the process that much more valuable with your kindness and eagerness for this project.

Executive Summary

This dissertation includes an outcome evaluation report of the LifeMatters Foundation Numeracy Programme. This programme focuses on strengthening the foundational numerical skills of the participants, in this case a group of Grade 2 learners from two schools in the Western Cape area. In total, these two schools had five Grade 2 classes of which constituted the sample. While this programme has run before, the LifeMatters Foundation decided to redesign the programme and run a new pilot programme in 2016. This dissertation focuses on the evaluation of this pilot programme with the goal to attain information on two outcome questions.

The first of these questions examined whether the programme participants' foundational numerical skills improved by the end of the programme and if they improved more than the skills of the comparison class. The comparison class for this evaluation was made up of 12 learners of one class that met the criteria for selection, but did not receive treatment. Each of the other four classes had the weakest 12 learners selected on the results of a class-based assessment delivered by the teachers. Therefore, in total, the evaluation included 60 participants. The second question examined if programme dosage, or the amount of attendance, was a significant contributor to the improvement of participants' numerical skills. As the programme was conducted over the course of the year, this question sought to control for the impact of maturation on the results and identify a programme effect.

Secondary data, provided by the LifeMatters Foundation, were used in order to answer the two evaluation questions. This data consisted of the results of the participants on eight measurements conducted throughout the year. These measurements were standardised tests, known as Formal Assessment Tasks, designed by the Western Cape Education Department. The data analysis methods included descriptive and inferential statistics for learners' performance and average programme dosage, a repeated measures ANOVA with a between-subjects factor for the differences between classes on each measurement, and a linear regression model for determining the effect of programme dosage on learners' final year mark.

List of Tables

Results highlighted that two of the four classes were significantly different from the comparison class. Furthermore, analysis revealed that, on average, the programme was not having the desired effect on the learners' performance. These results must be interpreted with caution as there was an issue of overcoverage in the programme. This refers to the ratio of participants in the programme that should not be in the programme over the total number of participants. More than half of the participants should not have been included in the programme, as they were far more academically advanced than the rest of the participants. In order to improve this facet, it is recommended that the LifeMatters' foundation develop a selection measure that is standardised, valid, and reliable.

The second evaluation question dealt with the impact of programme dosage on overall final mark, and as the average attendance of the programme was approximately 50%, there was no significant impact of attendance on final year mark. It is suggested that the requirements for attendance be re-evaluated as the low attendance rates played a role in the low programme effect.

The evaluation was limited by a lack of an adequate comparison of groups at baseline, as well as poorly controlling for maturation, a threat to internal validity, through the poor attendance. Despite the limitations, the evaluation has provided useful information for programme improvement, and if the recommendations are followed further evaluations will provide more conclusive results around programme effect.

Table of Figures

Table of Contents

Introduction

Research has shown that the mathematical ability of children in developing countries, such as South Africa, is behind that of its developed counterparts (Ramani, & Siegler, 2011). There are many common factors that contribute to this state of affairs. It is quite common for classes to be very large; often more than 35 learners are accommodated. Furthermore very few of these learners are taught in their home language (Setati & Adler, 2000). Adler and Setati (2000) also highlight the difficulty South African learners face when the language of teaching and learning (LOTL) in mathematics is not the language these learners are comfortable in.

While language is an important contributing factor to the aforementioned disparity, it is important to note that the South African context itself is characterised by disparity between schools. Research has indicated that South Africa has, in practice, two distinct schooling systems; that of the wealthy schools and the poorer schools (Spaull, 2013). Spaull (2013) highlights that there is a bimodality in results when measuring numeracy achievement in the historically white/wealthy schools and the historically disadvantage/poorer schools. This is echoed in similar research which indicates that the majority of South African primary schoolchildren finish primary school with a severe lack of numeracy knowledge and literacy (Fleisch, 2008). Therefore, the average results for numerical achievement in South Africa should be examined critically, as 25% of the primary school population who attend wealthier schools have vastly improved results in comparison with 75% of the primary school population who attend poorer schools. For instance, the 2014 Annual National Assessment Report on Foundation Phase Mathematics for Grade 1 and 2 learners showed that the overall performance of the sampled learners for those grades were 71% and 63% respectively (Department of Basic Education , 2014). However, the Systemic Results of Western Cape Schools showed that schools considered underprivileged produced much lower results in Grade 1 and 2 Mathematics than the mean and modal results. In a 2004 systematic evaluation of every primary school in the Western Cape, the Western Cape Education Department found that previously disadvantaged schools (formerly known as DET and HOR schools) showed significantly weaker results in Mathematics pass rates (Flesich, 2008). The following table has been extracted from these results:

Table 1

Numeracy Pass Rates by Grade and ex-Department

Ex-Department	Grade Three Numeracy Pass Rates	Grade Six Numeracy Pass rates
Model - C	90,5%	62,4%
Department of Education and Training	12,0%	0,1%
House of Representatives	30,5%	3,8%

Note. Adapted from WCED, 2004 as cited by "Primary Education in crisis: Why South African schoolchildren underachieve in reading and mathematics", Flesich, B, 2008, pp. 9 -10.

Further research, under the Monitoring Learning Achievement study, highlighted that South Africa performed, on average, the poorest amongst 12 African countries at a Grade Four numeracy level (Chinapah et al, 2000 as cited by Fleisch, 2008).

This disparity, and the difficulty learners in rural and urban schools face when learning mathematics, highlights a clear need for numeracy interventions and programmes targeted at improving the mathematical performance of primary school learners in these schools. These learners are only acquiring the basic and rudimentary skills, knowledge, and concepts in the field of mathematics, resulting in limited mathematical ability that sets the learner up for future difficulty (Flesich, 2008). It has been shown that early mathematical interventions have been a significant factor in improving future academic success in the field (Aguilar, Marchena, Menacho, Navarro, Ruiz, & Van Luit, 2012; Barnett, 1995; Claessens, et al., 2012). Essentially, the difficulties in mathematical performance for a learner is often the result of having limited experience and exposure in numeracy education, and therefore numerical interventions provide a persistent effect when the intervention is designed and implemented correctly (Siegler & Ramani, 2011a). Ramani and Siegler (2011a) note that increasing the number of young learners receiving mathematical interventions, as long as they are grounded in sound programme theory, is a worth-while goal.

The LifeMatters Foundation, a not-for-profit organisation, supports these beliefs and is one of many organisations committed to improving the numeracy performance of learners from underprivileged schools in Cape Town. This dissertation aims to provide a formative evaluation of LifeMatters Foundation's first implementation of a modified numeracy programme in two schools in the Western Cape Province. A detailed description of this intervention follows.

Programme Description

The following programme description was compiled from various documents and reports provided by the LifeMatters Foundation (LMF) for the purposes of this evaluation. They included annual reports of the foundation (2015; 2016), the strategic plan (2016), volunteer forms (n.d), brochures (n.d), and their website (www.lifemattersfoundation.org).

The LifeMatters Foundation (LMF) is a Christian Non-Profit Organisation (NPO) that was founded in 2002 by Meadowridge Baptist Church, now known as Connect Church Meadowridge. The aim behind this development was to serve the youth of the Southern Suburbs and surrounding areas in the areas of education, life skills, and emotional support. The LMF is headed by an Executive Director who oversees the administration and running of the operations. LMF is supported by a number of donors such as the Claremont Rotary, Newlands Rotary, Chic Mamas Do Care, as well as Connect Church Meadowridge. The LMF Operations Manager is responsible for the fundraising of the organisation, and they hold an Annual Fundraising Event.

Since 2002 LMF has been involved in 11 primary schools across their four projects: Literacy, Numeracy, Life Skills, and Counselling. Two of these projects, namely the Numeracy and Literacy projects, constitute what is known as the Academic Portfolio for the LMF, which is headed up by the Academic Portfolio Manager. Currently these projects are implemented in four schools in the Retreat and Steenberg areas. The LMF reports that the dropout rate that these schools see at the end of Grade 7 is 50%.

Based on the idea that improving the performance of learners in numeracy at a young age will result in a decrease in learner dropout and improved performance in the long-term, the LMF seek to expose school children to the help they require to reach their full academic potential, with the hope that children will stay in school and succeed. The LMF identified these schools as those which are heavily affected by poverty and the lack of educational support that is representative of the education crisis. This information informs the high level goals and objectives that the LMF seek to achieve:

- To empower children to make right responses in the situations in which they find themselves.
- To support and encourage children towards reaching their full academic potential.
- To develop the sense of purpose and self-worth of children within a healthy value system.
- To equip adults (parents, teachers, and volunteers) in their various roles with children.

Each of the LMF's goals are addressed by one of their four projects, and therefore the Numeracy Programme seeks to achieve the second goal in the above list. In addition to the organisational objectives, the Numeracy Programme "aims to improve poor numeracy results by strengthening learners' foundational skills for Numeracy while they are in Grade 2" (LMF, 2015). The primary means of achieving this goal is through the work of the volunteers and the Numeracy Coordinator.

In 2014 the Numeracy Programme assessed Grade 2 learners, and the weakest 20 children in each class were selected to participate in the programme. From there trained volunteers (who were recruited and trained by the LMF) sat with two learners each twice a week, and exposed them to various tasks and games that aimed to improve their basic number concepts and key addition and subtraction skills. While this programme showed that learners were improving, due to the inconsistent nature of the performance and a small sample size, there was limited quantitative analysis.

In 2015, the LMF decided to implement the programme through the use of a resource and education manual titled "My Fun With Numbers". The manual was based on peer education, and a Grade 6 learner would be paired with a Grade 1 learner and, under the supervision of the LMF volunteers, would work through the manual. However, this came with its own set of challenges regarding attendance and the mathematical skills of the grade 6 learners. Therefore, in 2016, the LMF decided stop the peer education element.

In 2016 the programme was implemented in two of the four schools discussed above, and the sample of participants was made up of the Grade 2 learners in those schools. For ethical reasons, these schools are referred to as School A and School B.

The original plan was to use the Annual National Assessments (ANAs) results from the learners Grade 1 year to select learners for the programme. However, because the ANAs were cancelled in 2015, the programme had to improvise and select learners based on their results on assessments developed by the school teacher. As a result, two separate assessments where administered at each school in order to select the participants of the programme. As the two schools had an uneven number of classes, with School A having three Grade 2 classes and School B having two Grade 2 classes, one of the classes in School A was selected to be a comparison group. In the end, 12 learners from each class were selected for the programme on the basis that they were the 12 weakest performing learners in each class. While the comparison class was made up of 12 learners that met the criteria for selection, they received no treatment. The programme ran twice a week with volunteers working with two learners each for half an hour. LMF believed that these small groups were more effective for the learners, as they often came from classes of 60 learners where the teacher did not have time to provide learners with individual attention. Each school has designated classrooms to be used for LMF's numeracy and literacy programmes. The Numeracy Coordinator supervised the volunteers in order to ensure that they are doing their job correctly and according to the guidelines set by LMF. Each volunteer was required to attend training workshops facilitated by the numeracy coordinator.

In the numeracy programme the volunteers facilitate weekly activities described in the "My Fun with Numbers" book. The 30-minute sessions were broken up into three parts. The initial five minutes were spent on a "Speed Writing" activity that focused on improving the learner's counting skills. The majority of the time (around 20 minutes) was spent on the main activities described in the book. The "My Fun with Numbers" book was designed by an occupational therapist who worked with LMF in 2014 and 2015, and used the Curriculum and Assessment Policy Statement (CAPS) set of objectives for Grade 1 mathematics as a baseline for design. The CAPS (Department of Education, 2011) focus areas of content for Grade 1 mathematics are seen as follows in Table 1:

Table 2.

Weighting of Content Areas in Grade 1

Content Area	Weighting of Content Area for Grade 1
Numbers, Operations, and Relationships	65%
Patterns, Functions, and Algebra	10%
Space and Shape (Geometry)	11%
Measurement	9%
Data Handling (Statistics)	5%

Note. Adapted from "National Curriculum Statement: Curriculum and Assessment Policy Statement. Foundation Phase," by the Department of Education, 2011, p.15.

The reason that Numbers, Operations, and Relationships is the majority of the weighting for Grade 1 is because CAPS states that:

> Learners need to exit the Foundation Phase with a secure number sense and operational fluency. The aim is for learners to be competent and confident with numbers and calculations. For this reason the notional time allocated to Numbers, Operations, and Relationships has been increased. Most of the work on patterns should focus on number patterns to consolidate learners' number ability further. (Department of Education, 2011, p. 10).

This weighting is also found in the assessments used by the Western Cape Education Department for their standardised assessment mark allocation as the percentages seen in Table 2 represent what each content area is allocated for scoring. The LMF Numeracy programme attempts to replicate this weighting through the design of the intervention, as they were designed primarily to improve the learners' outcomes with this area, however the other focus areas were also included throughout. Each of these content areas and how they relate to programme activities are discussed in the following sections below.

Numbers, Operations, and Relationships.

This content area refers to the performance of a learner on tasks related to number sense, which involves understanding the meaning of different numbers, the relationship between these numbers, the difference in size between one number and another, the representation of a number in different ways, and the effect of operating with numbers (Department of Education, 2011). While this is reflected in the South African Foundation Phase curriculum, it is also echoed in other parts of the world. For instance, the Institute of Education Sciences published a report titled *Teaching Math to Young Children* and one of the primary recommendations for teaching mathematics to learners of a young age is a developmental and incremental focus on numbers and operations (Baroody et al, 2013). The Numeracy Programme uses these concepts as the foundation of the programme activities. These games and activities incorporated physical objects and games that aimed to improve performance in counting, number recognition, number identification, number sense, and problem solving.

Patterns, Functions, and Algebra

The importance of algebra is well-documented in both national and international curricula (Baker et al, 2014; Baroody et al, 2013; Department of Education, 2011). It is considered the language and means of communication in Mathematics, and in the Foundation Phase there is primary focus on number patterns and geometric patterns (Department of Education, 2011). The activities and games conducted within this content area made use of physical objects to create patterns, whereby the learners can follow and copy the pattern so that they develop the skills required to see the logic behind a pattern. Importantly, this content area supplements the concept development and operational sense that is developed in the Numbers, Operations, and Relationships.

Space and Shape (Geometry)

In this content area learners learn to recognise, describe, and create shapes and objects. The programme activities involved games and activities that attempted to strengthen the learner's performance in these areas through drawing shapes and describing objects they see and interact with over the course of the sessions. Importantly, the sessions sought to improve the learners understanding of position, such as left and right, on top of, or upside down, with the aim of recognising and matching everyday objects from a difference perspective or orientation.

Measurement

The programme activities sought to improve the learners performance of measuring units such as kilograms, centimetres, and litres. The games and activities used everyday objects such as bottles and food items to show how these properties can be measured and described. The sessions were also timed, and this was communicated to the learners so that their concept of time in regards to measurement could be improved.

Data Handling

While one of the focus areas stipulated in CAPS, this was not focused on in the Numeracy Programme.

Session Activities

Every week the volunteers facilitated three activities over the two sessions, which would focus on one or more of these content areas. In this way, learners had focused time to improve the concepts in an interactive and supportive way, with the hope that the increased attention from small group settings and the entertaining way the games and activities are designed would be the driving force of this improvement. The final part of the session, if there is time remaining, involved revision and helping the learners with any questions or queries.

In 2016, the entire programme took place over eight months starting at the beginning of April and ending in December, where learners were tested for improvement eight times during the year: twice in each term. The implementation of the programme can be illustrated by means of a service delivery and utilisation flowchart (see Figure 1).

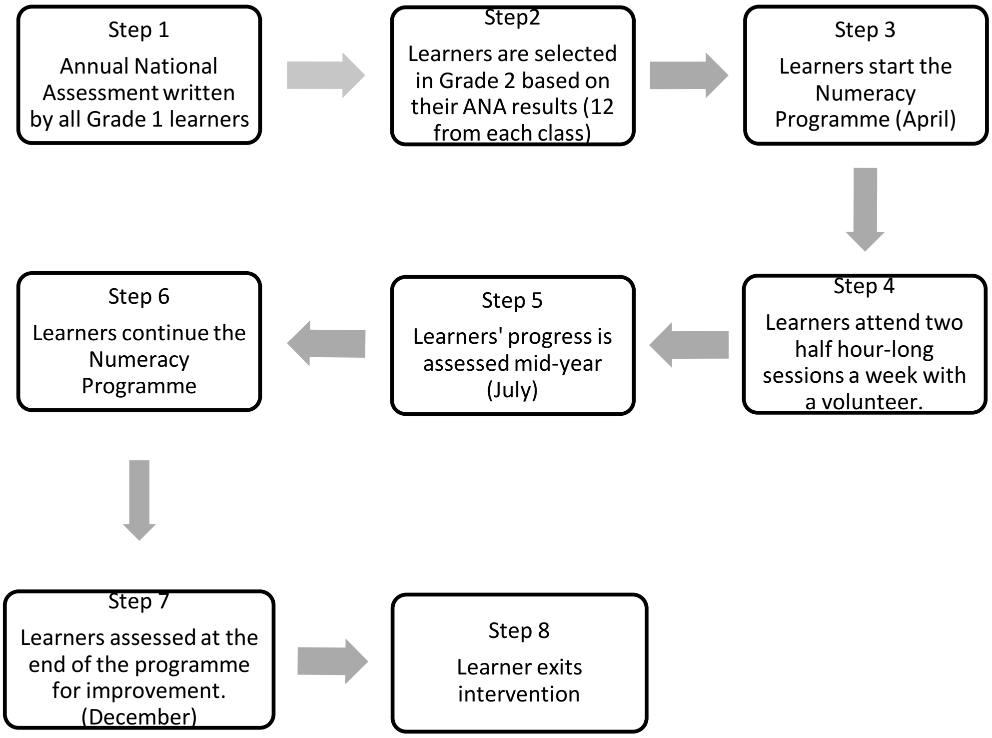

Figure 1. The numeracy intervention's service delivery and service utilisation flowchart for 2016.

As the LMF seeks to provide a positive space in which a learner receives the support and attention required in order to reach full academic potential, it is valuable to assess whether the numeracy programme is able to bring about the change it has been designed to affect. In the following section the focus shifts towards how this change will come about and this is depicted and discussed as a programme theory, or a theory of change.

Programme Theory

Rossi, Lipsey, and Freeman (2004) describe programme theory as a foundational concept in the process of developing evaluation questions, designing an evaluation, and, ultimately, discussing the findings of an evaluation. Essentially, the programme theory of an intervention states that if the right resources are provided for the right activity presented to the right people, then the results of the intervention will achieve the outcomes set for the programme (Jordan & McLaughlin, 2010). This concept is what Rossi, Lipsey, and Freeman (2004) refer to as impact theory. LMF does not have an articulated programme theory, and therefore it can be considered as implicit (Rossi, Lipsey, & Freeman, 2004).

For the purpose of this evaluation, the programme theory for the Numeracy Programme was elicited through the information gathered in meetings between the Academic Portfolio Manager, the Programme Director, and Numeracy Co-ordinator as well as various pieces of documentation provided. Figure 2 depicts the LMF Numeracy Programme's programme theory as a logic model. A logic model is a tool, often depicted in a figure or diagram, that provides insight into the processes and activities that lead to the various outcomes of a programme or intervention (Padgett, Royse, & Thyer, 2010). The Academic Portfolio manager confirmed that the programme theory shown in logic model format, is an accurate representation of the Numeracy Programme.

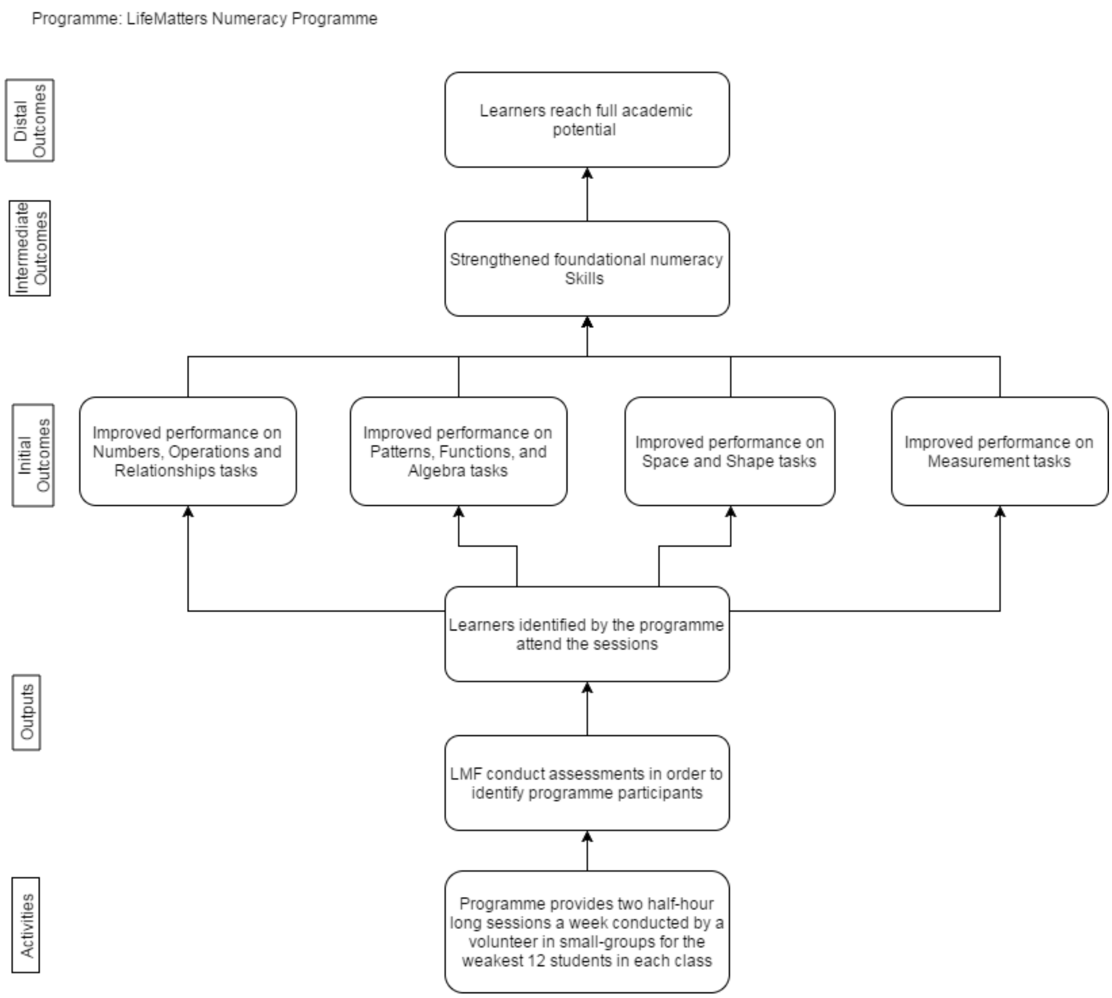

Figure 2. Logic Model of LifeMatters Numeracy Programme. Adapted from, "Evaluation: A Systematic Approach", Rossi, Lipsey, & Freeman, 2004, p.95.

Plausibility of Programme Theory

While the logic model identifies the resources and activities that lead to the Numeracy Programmes three core outcomes, it does not explicitly state the underlying assumptions the programme makes. The numeracy programme operates under three causal assumptions, namely:

- Small group sessions are effective in numeracy interventions
- The activities in the intervention will strengthen foundational numeracy skills
- Strengthening foundational numeracy skills will help students reach their full academic potential in the future.

In order to examine the plausibility of these assumptions, a literature review of early developmental numeracy programmes and their intervention strategies was conducted. As a literature review attempts to answer a review question, each of the assumptions above was moulded into a question. The review was conducted between the 9th of March 2016 and the 4th of April 2016, with the use of the online databases provided by the University of Cape Town's library. The search was conducted within two platforms, Google Scholar and EBSCOHost, using all available databases within those platforms. In order for an article to be selected for the review, it had to meet the following inclusion criteria:

- Written in English
- Full-text articles
- Peer-reviewed

In addition to the inclusion criteria, articles were excluded if they were published outside of the 2005 - 2016 date range. The search terms used to create search strings varied depending on the examined assumption (such as "mathematics intervention" AND grade 1" AND long-term effects, or, numeracy interventions AND small group AND effective). In addition to the keyword search, reference list mining was used to find relevant articles in the reference lists of appropriate articles. Articles found through this method that fell outside the time-frame mentioned above were used if they contained important or relevant information. Articles were then selected by title, and then by abstract in terms of their relevance to the review questions. At the end of this process, 16 articles were selected for the three review questions.

Are small group sessions effective in numeracy interventions?

As stated in the Programme Description, the LMF organises the Numeracy Programme sessions into small groups, with a volunteer working alongside two learners. The LMF believe that this provides the attention and support that these students lack when learning mathematics in classes of large numbers. However, an article by Howie (2003) stated that there was no significant relationship between class size and mathematical success, but rather the significant factor was the learners' efficiency in the language of learning. Therefore, before small group sessions and their effectiveness are examined, it should be vital that the LMF recruit and seek out volunteers who are comfortable in

Afrikaans (the primary home language of the targeted schools) as the effectiveness of the intervention is tied to language.

Cirino et al. (2008) identified seven principles that were valuable when assessing the effectiveness of an early mathematics intervention. One of these principles, instructional design, indicated that in order to eliminate and overcome misunderstandings and learning barriers in mathematics the means of instruction and delivery of the intervention are vital (Cirino et al., 2008). Cirino et al. (2008) have indicated that this principle was one of the most overlooked aspects of mathematics education, and that there should be a greater focus on individualising interventions in order to meet the participants where they are in relation to their ability. Small group interventions were one of the multiple suggestions when contemplating the implementation of instructional design.

A similar study, which primarily focused on mathematic interventions for first- and second-grade students (the target population of the LMF Numeracy Programme), found small-group instruction to be a recommended and effective means of instruction (Byrant, Bryant, Chavez, Gersten, & Scammacca, 2008a). However, the same study noted that Tier 1, or whole-class interventions, have shown success when dealing with low-achieving learners. Byrant et al (2008a) propose in their study that Tier 2 (small-group or flexible grouping) interventions are an effective instructional design that can produce significant results, and the findings of the study indicated that the intervention effect was significant for Grade 2 learners. However, the intervention effect was not found to be significant with Grade 1 learners. In a follow-up study the authors found that, with increased duration of the study and more sessions in the intervention, the intervention effect was significant for Tier 2 interventions at a first-grade level (Byrant, et al., 2008). The results of this study echoed that of others in the field, and there is evidence that small-group interventions are an effective, and necessary, means of instruction in first- and second-grade mathematic interventions (Byrant, Bryant, Gersten, Scammacca, & Chavez, 2008; Byrant, et al., 2008b; Bryant, Compton, Fuchs, Fuchs, Hamlett, & Paulsen, 2005; Cirino, Fletcher , Fuchs, Fuchs, Powell, & Seethaler, 2008; Fuchs, Fuchs, & Hollenbeck, 2007). Furthermore, the What Works Clearinghouse standards, a measure of an interventions causal validity and quality of practices, highlighted there is evidence that Tier 2 interventions are effective, however it is not possible to identify whether the impact of these interventions was solely due to the small-group or whether it was the combination of the small-group as well as the Tier 1 (whole-class) instruction (Beckmann et al, 2009). With that being said, there is enough evidence supporting the theory behind the use of small-group instruction in Grade 2 numeracy programmes.

Is there evidence that early mathematics interventions will strengthen foundational numeracy skills?

At the heart of the Numeracy Programme there is the goal of improving the mathematical performance of struggling learners. Research has shown that young children have an innate ability to learn mathematics far beyond what they are exposed to in school (Clements & Sarama, 2011). While this may be the case, many children across the world have not been able to fulfil this potential as they lack the exposure, experience, and quality education necessary, and this is certainly the case in the South African context. Clements and Sarama (2011) argued that without high quality education, children, especially those from rural or underprivileged contexts, will not succeed in the field of mathematics. The authors noted that "research-based interventions...positively affect children's competencies in mathematics" (Clements & Sarama, 2011, p. 696). Ramani and Siegler (2011b) showed that even though children from low-income backgrounds have a weaker knowledge of numeracy, when these children participated in an intervention focused on playing numerical based board games designed to increase foundational numerical skills their numerical skills improved to the point where they were indistinguishable from upper to middle class learners who did not play the games. This research provided an example of the causal evidence of numerical interventions strengthening the foundational numeracy skills of the participants. Similarly, a review conducted on 10 interventions across 5 different countries found that there was increasing neurological and cognitive evidence for the effectiveness of both short- and long-term mathematics interventions (Dowker, Heine, Kadosh, Kaufmann, & Kucian, 2013).

The research cited here provides plausible evidence indicating that numeracy based interventions strengthen the foundational numeracy skills of participants. Some of the reviewed interventions made use of numerical board games, or activities presented in an engaging and entertaining manner, rather than what is presented in a classroom setting, and this proved to be effective. The CAPS report briefly mentions this in the section entitled *Learners with barriers to learning Mathematics* stating, "It is important for learners who experience barriers to learning Mathematics to be exposed to activity-based learning. Practical examples using concrete objects together with practical activities should be used for a longer time than with other learners" (Department of Education, 2011, p. 12). Although this is referring primarily to classroom-based activities, it indicates that since the Numeracy Programme sessions are presented in a similar manner, that the theory behind this intervention is sound.

What causes the effects of a mathematics intervention to persist?

Of the three review questions, this question yielded mixed findings. There was a consistent finding within selected articles that a lack of basic numerical ability during the primary school years of the learner influences their ability to acquire more advanced mathematical techniques (Aguilar, Marchena, Menacho, Navarro, Ruiz, & Van Luit, 2012; Byrant, Bryant, Gersten, Scammacca, & Chavez, 2008a; Davis-Kean, Duncan, Siegler, & Watts, 2014; Toll & Van Luit, 2012), however the results concerning the long-term effects of early numerical interventions appear to be mixed.

Bailey et al. (2015) found that although the post-test scores of the learners from 42 low-resource schools who participated in Building Blocks, an American early mathematics intervention, were higher than the learners who were part of the control group, the effect attenuated over time but remained statistically significant in follow-up assessments one year later. Fadeout can be understood as the rapid decline of the effects of an educational intervention, and, if not attended to, can influence the effect the intervention has on future academic success (Bailey, et al., 2015). In order to combat fadeout, the authors noted that the effects of an educational intervention, such as an early mathematics intervention, will be sustained as long as the content provided in school becomes more advanced after the intervention ends. It is easy to see the problem this presents in the South African context, where the bimodality of the South African schooling system exists and the quality of education in underprivileged schools is lacking. Many schools teach mathematics to the early grades (Foundational Phase) at the level most appropriate for the weakest students in the class (Bailey, et al., 2015; Eksteen, 2014), and while this is often necessary it does not prevent the effects of the intervention from fading out. Once the intervention ends, and the learners are back in the classroom, they no longer receive the same level of cognitive and educational stimulation provided by the intervention and tend to regress towards the mean.

However, there seems to be evidence that shows that fadeout and the persistent effect of the intervention is affected by the quality of the intervention's design (Adelstein, Gormley, & Hill, 2012; Barnett, 2011). This is consistent with the earlier study of Cirino et al. (2008) where instructional design was one of the seven principles found to correlate with effective mathematic interventions. The review found that when an intervention was research-based and designed in such a way that it was supported by developmental and cognitive theory, the long-term effects of early childhood interventions produced sizeable effects on grade retention and even high-school academic success (Barnett, 2011; Barnett, 1995; Davis-Kean, Duncan, Siegler, & Watts, 2014). In conclusion, the lasting effects of the Numeracy Programme is tied to the quality of the design as well as the instruction provided to the learners. Since it appears plausible that the design of the programme is of sufficient quality based on the literature, the instruction of the learners is required to be at the same standard to achieve lasting effects that would fulfil the distal outcomes of the programme.

From the literature reviewed here to address the three review questions, helpful insights into the plausibility of the LMF's Numeracy Programme theory were gained. The selected articles highlighted that small-group interventions are effective in terms of intervention delivery, and that although there is a paucity of research around first- and second-grade mathematics interventions as the focus tends to be on higher-grade learners, there is enough to validate the LMF's small-group design. In addition, there is plausible evidence that more advanced interventions strengthen numeracy skills and prevents fadeout.

Evaluation Questions

This year (2016) is a pilot year for a new version of the LMF’s Numeracy Programme. It would seem that this revised programme is grounded in a plausible programme theory. In order to assess whether this revision works, the proposed evaluation will be an outcome evaluation. An outcome evaluation seeks to assess whether a programme has changed the state of affairs or improved the lives of its beneficiaries (Rossi et al.,

2004). In the proposed evaluation, this means a change in mathematical performance for those learners who attended the programme.
Based on the logic model depicted in Figure 2, one evaluation question has been formulated in cooperation with the programme manager:

1. **Are the foundational numerical skills of the participants of the Numeracy Programme better after the programme than before, and better than those of learners who did not take part in the programme?**

In the light of the poor attendance rate for the 2015 programme and in order to analyse the outcome results in a more nuanced way, the researcher has added a second evaluation question:

2. **Is there evidence that learners who attended more of the programme were better off in terms of mathematical performance than those who attended less of the programme?**

The method used for answering these questions is described in the following section. This evaluation will be using a theory-driven evaluation science approach as described by Donaldson (2007)

Method

An important aspect of evaluation research is designing an evaluation that can be repeatable and have a real-world effect. The following sections describe, in detail, the design and method of the evaluation.

Design

A quasi-experimental non-equivalent group design was employed to explore whether the foundational numeracy skills of the participants of the Numeracy Programme was better after the programme than before and better than those who did not take part in the programme. A quasi-experimental design was specifically chosen, as there was no random assignment of participants to either the treatment or comparison group. While an experimental design would have been more effective at attributing a causal relationship between the programme and the measured effect, in evaluation research it is often not possible to meet the requirements of an experimental design (Babbie, 2013). A strength of the non-equivalent group design is that is still controls for various threats to internal validity, albeit it not as effectively as an experimental design.

As a result, the design included five groups, a set of four groups who participated in the intervention, and a comparison group of learners who did not participate in the intervention. These groups were divided by class, three classes from School A and two classes from School B, with the comparison group being from class 2B from School A. Thus, the design could analyse the groups by class (5 groups, 4 treatment and 1 comparison) and the information provided was used to determine whether or not the programme was effective in improving the mathematical ability of the participants.

Additionally, a descriptive design was used to explore whether or not programme dosage (in other words, how much of the programme the learners attended) played an important role in the mathematical performance of the participants of the programme, as some learners participated more than others. The results of the four intervention groups were used in this design.

Participants

Participants were selected for the programme based on their scores on a single mathematical assessment prior to the intervention. These assessments were created by the teachers at the two schools and then administered to all Grade 2 learners in those schools. Each school administered a separate assessment to the Grade 2 learners, and the results of these two separate assessments were used to select the weakest 12 students in each class. While the initial plan was to use the Annual National Assessment (ANA) results from the Grade 1 numeracy test as it would have provided an adequate comparison and standardised result between all learners, it was not possible due to the postponement of the 2015 ANAs to early 2016 by the Department of Basic Education. This delay caused disruption amongst schools, and ultimately the ANAs were not conducted in every school as they should have been, and the schools that included in the programme were amongst those schools which did not administer the assessments. As a result, the schools designed these separate assessments as class-based tasks so that there was at least some means of selection. The consequence of this will be discussed in later sections.

The treatment and comparison groups consisted of Grade 2 learners from two primary schools. For ethical reasons, the schools will not be identified and will simply be referred to School A and B.

The comparison group included a group of 12 learners from one Grade 2 class at School A who met the eligibility criteria but did not participate in the programme. However, two learners did not participate in the programme at all during the year, and thus were treated as non-participants for the purposes of this study. One of these learners were in School A and class 2C, and the other in School B, class 2B. It must be noted that School B selected one of the classes (2A) on the basis of academic merit, and therefore that may influence programme outcomes. However, School A's classes were made up of students of various academic ability, according to the Head of Grade for this school.

Table 3 depicts the population of Grade 2 learners, and the assignment to either the treatment or comparison group.

Table 3

Study Population and Sample (n = 60)

School	Class	Number of Total Learners	Learners eligible	Treatment or Comparison Group
School A	2A	34	13	Treatment
	2B	33	12	Comparison
	2C	34	11	Treatment
School B	2A	39	13	Treatment
	2B	38	11	Treatment
TOTAL:		178	60	

A separate teacher taught each class, and the five teachers worked in conjunction with the Numeracy Co-ordinator, the Principals of the two schools, and the volunteers to facilitate the programme and monitor the progress of each participant throughout the year. The volunteers are members from the local community that volunteer as a means of community service and goodwill. These volunteers work under the supervision and training of the Numeracy Programme Co-ordinator.

The learners were selected for the programme during the first term, and completed the programme at the end of 26 weeks, which finished at the end of the final term in December. During that time the participant's performance was measured across eight separate measures, which will be discussed below.

Measures and Procedure

Secondary data, as provided by LMF, was used as part of the evaluation. The data consisted of learners' performance measures, namely the scores achieved on eight measures: two assessments administered to all learners during the course of each term during the academic year. This was because the nature of the Mathematics subject in Grade 2 included two Formal Assessment Tasks (FAT) each term that made up the overall term mark (50% per FAT). Therefore, it was possible to analyse at the measurement level, as each FAT assessment covered the same five activities throughout the year, or at the overall term mark level. These five activities are listed below, with the mark allocation in parentheses:

1. Numbers, Operations, and Relationships (65%)
2. Patterns, Functions, and Algebra (10%)
3. Space and Shape (11%)
4. Measurement (9%)
5. Data Handling (5%)

Each FAT assessment that the learners wrote consisted of questions revolving around each of these five activities. As the learners were expected to progress throughout the year in these areas, the Numeracy Programme aimed to equip learners in four of these five areas, with the exception being data handling, as that was not explicitly dealt with by programme activities.

In each assessment the learners were assessed in terms of their development in the five content areas. These assessments were developed by the Western Cape Education Department, and designed to test learners on the five aforementioned activities fairly, and the standardised nature of these tests were useful in the comparison of results between and within groups. The tests are out of 50 marks, and for the purpose of analysis these marks were converted to a percentage score. Each term the learners wrote two of these tests that comprised their final term mark, and the final year mark consisted of the average across all eight assessments. For the purposes of this study, Term 1, 2, 3, and 4 results were used as data points for each learner, as well as the final year mark as an additional source of information. In addition to the marks, LMF kept an attendance register for each week of the programme, and therefore this data was used as a source of information around programme dosage.

Ethics

Permission to conduct the evaluation was granted by LMF's Programme Director (see letter of permission in Appendix). In addition to this permission, ethics approval and clearance to make use of the secondary data was provided by the Ethics in Research Committee of the Commerce Faculty at the University of Cape Town. It is important to note, that because of ethical requirements, the students that were excluded from the pilot programme were scheduled to get the programme at a later stage.

Data Analysis

In order to analyse the data for the purpose of the evaluation, descriptive statistics such as means and standard deviations, were used along with inferential statistics such as a repeated measures Split-Plot ANOVA to measure the comparison between groups. Each of the eight measures were testing the same concepts and therefore can be compared across time using the repeated measures ANOVA with a between-subjects factor being the group of the participant. The mean results for each assessment were analysed using graphical tools, and compared to a cut-off point. That cut-off point was 40%, the standard expected of Grade 2 learners in Mathematics for progression. In addition to these measures, regression analysis was used to analyse whether group placement was a significant predictor of final year mathematics scores, and to analyse whether attendance in the programme was a significant predictor of Mathematics scores. All quantitative data was analysed using IBM SPSS Statistics Version 23, and the graphical representations were produced using SPSS and Microsoft Excel.

Results

The results are presented in line with the evaluation questions that were presented at the end of the method section.

1. **Are the foundational numerical skills of the participants of the Numeracy Programme better after the programme than before, and better than those of learners who did not take part in the programme?**

In order to answer this question a mixed-design ANOVA, or Split-Plot ANOVA, was conducted. This method was used as there were repeated measures on the outcome variable, as well as a between-subjects factor, namely the class of the learner. There were two steps in this analysis, first the analysis compared the results on each of the eight assessments against the five class groups, and secondly the results on final year marks against the five class groups. For all statistical reporting below, the alpha level used for significance testing was $p < 0.05$.

Two assumptions were initially checked before the analysis was run, namely for normal distribution of the data and homogeneity of variance. Histograms and explorative descriptive statistics indicated that the data was normally distributed. In the Table below (Table 4), the descriptive statistics for each group can be found, with the comparison group statistics indicated by a different colour. Each of the two assessments per term (eight in total) are reflected here.

Table 4.

Descriptive Statistics for classes in each school.

Assessment	Class of Learner	Mean	Std. Deviation
Score of learner on Assessment 1	School A, 2A	57.85	9.34
	School A, 2B	58.00	13.92
	School A, 2C	43.09	12.83
	School B, 2A	48.46	6.62
	School B, 2B	45.45	19.36
	Total	50.87	13.39
Score of Learner on Assessment 2	School A, 2A	64.08	11.50
	School A, 2B	37.58	12.82
	School A, 2C	32.72	13.58
	School B, 2A	35.23	9.61
	School B, 2B	32.55	12.61
	Total	41	16.96
Score of Learner on Assessment 3	School A, 2A	72.77	8.43
	School A, 2B	25.92	13.60
	School A, 2C	32.90	13.79
	School B, 2A	43.70	12.65
	School B, 2B	36.45	15.73
	Total	43.13	20.91
Score of Learner on Assessment 4	School A, 2A	51.77	15.80
	School A, 2B	30.50	13.24
	School A, 2C	27.09	12.13
	School B, 2A	58.00	8.35
	School B, 2B	34.63	10.96
	Total	41.20	17.27
Score of Learner on Assessment 5	School A, 2A	72.08	24.97
	School A, 2B	36.67	13.12
	School A, 2C	31.27	13.52
	School B, 2A	62.92	15.57
	School B, 2B	38.90	15.03
	Total	49.45	23.36
Score of Learner on Assessment 6	School A, 2A	78.77	10.70
	School A, 2B	50.25	12.46
	School A, 2C	48.45	15.90
	School B, 2A	67.85	11.87
	School B, 2B	49.27	11.77
	Total	59.73	17.44
Score of Learner on Assessment 7	School A, 2A	70.85	14.85
	School A, 2B	47.50	16.62
	School A, 2C	33.54	17.65
	School B, 2A	52.62	14.99
	School B, 2B	46.18	12.56
	Total		
Score of Learner on Assessment 8	School A, 2A	78.84	18.22
	School A, 2B	47.83	23.17
	School A, 2C	32.81	17.24
	School B, 2A	62.54	19.59
	School B, 2B	40.18	12.24
	Total		

* Comparison group's performance in blue.

This table can also be presented as a clustered column graph (Figure 3), and a line graph (Figure 4) below. The comparison group is indicated with a dark blue bar in both figures. The horizontal line in the figures indicates the pass mark for mathematics in this grade.

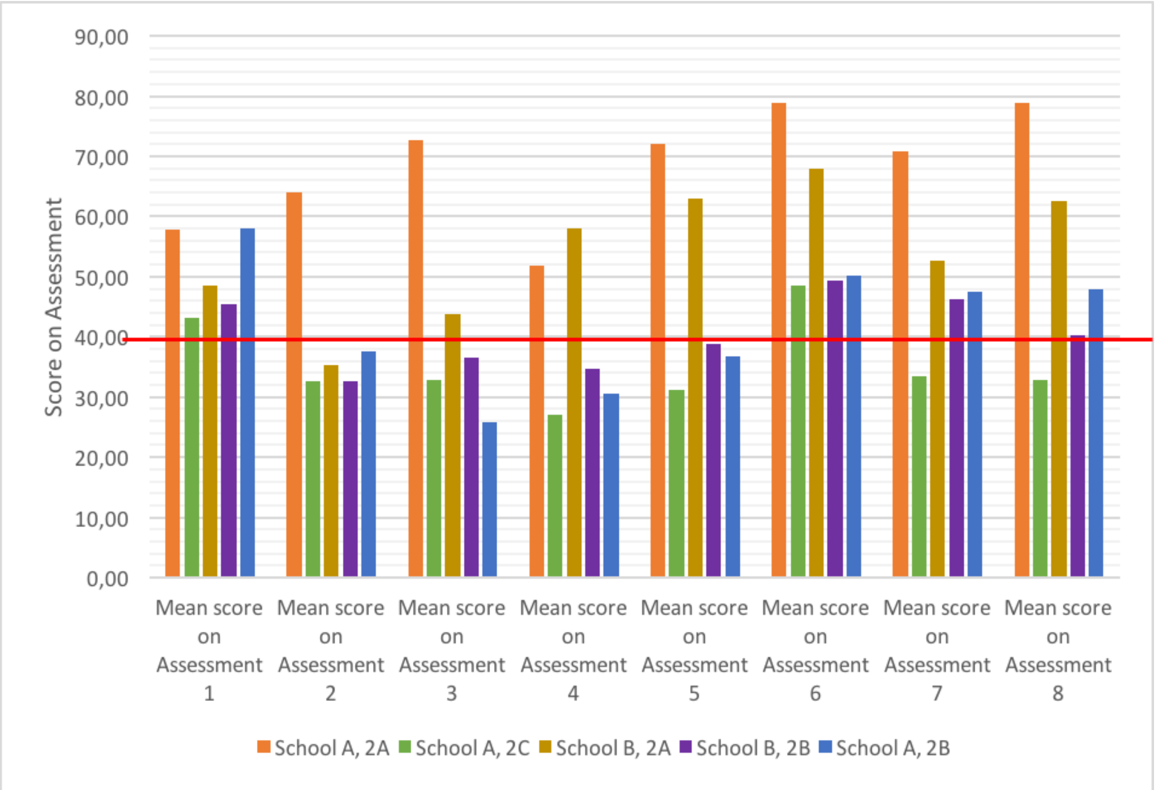

Figure 3. Average Marks on FAT Assessments 1 – 8 grouped by Class and School.

As there were eight assessments across the academic year, this graph is useful when comparing the average results of each class as well as the results against the 40% mark which is the progression requirement for Grade 2 learners under the CAPS administration for 2016 (Department of Basic Education, 2012). While these results will be discussed further in the following section, it is interesting to note the steady progression in the scores of School A, Class 2A and School B, Class 2A.

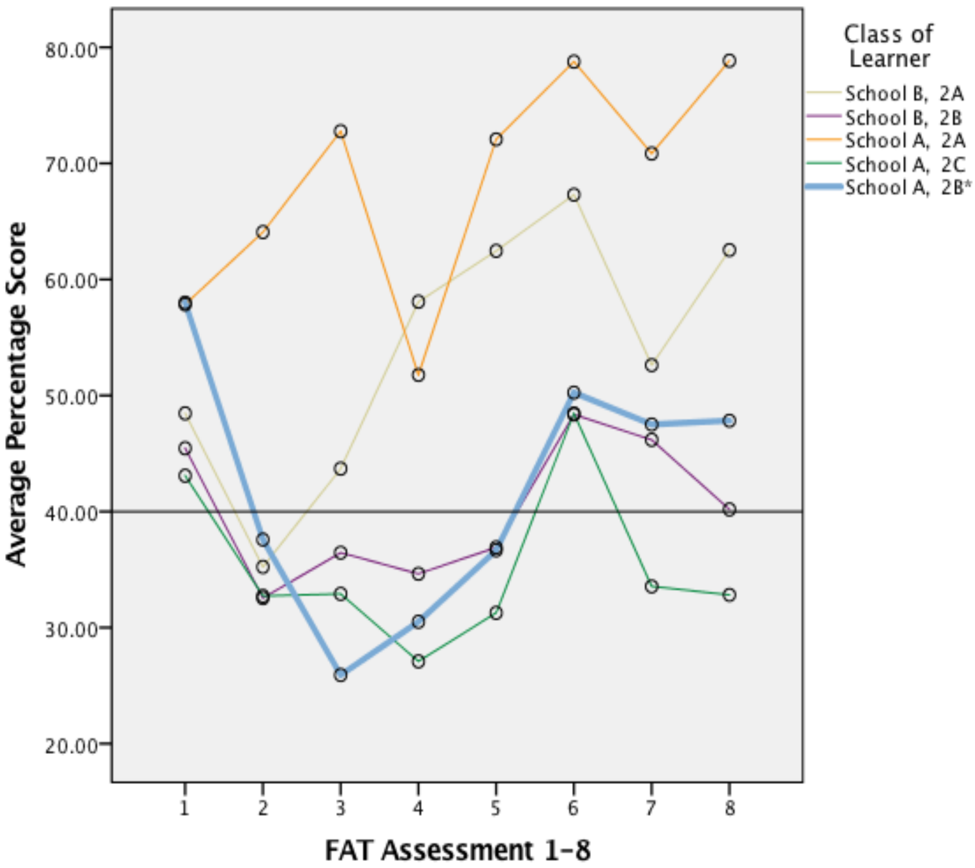

Figure 4. Means of Class against Assessments measured.

In a similar fashion, this Figure presents the progression of each class, and while two classes (School A, 2A and School B, 2A) were consistently higher than the comparison class (indicated by the darker blue line), the other classes results seem to be similar to the comparison class.

Another interesting comparising can be seen in Figure 5 below, as it compares the year average mark (which is the final year mark for the learner) for each learner included in the analysis.

Only one class, School A, 2A (indicated by the orange columns), had a 100% of the participants above the threshold for progression.

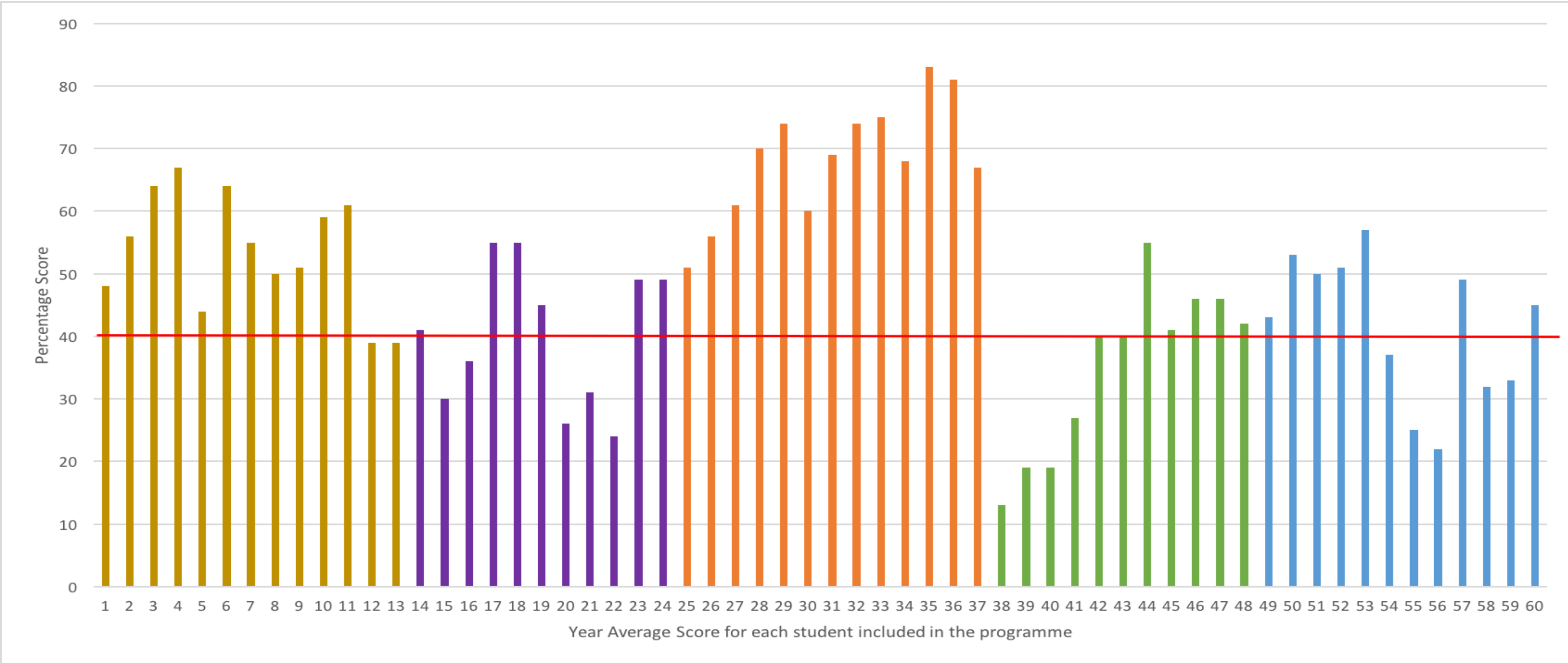

Figure 5. Year average score for each learner included in the programme.

**Note.* Classes are indicated by the various colour lines, and the comparison class indicated by the light blue bar, (49 – 60).

From Figures 3 – 5 it is clear that classes A in School A and B outperformed all other classes in all assessments. In the following section, more in-depth statistical analyses are used to determine whether this difference was statistically significant.

Once the assumptions of normality and homogeneity of variance had been checked and were not violated, a third assumption, the assumption of sphericity was tested. ANOVAs with repeated measures are particularly at risk of violating the assumption of sphericity, a condition that involves the variances of difference between all the combinations of the groups included in the analysis and is of similar importance to the homogeneity of variance assumption for between-subjects ANOVAs (Field, 2012). If the assumption is violated, then there is a higher risk of a Type II error as the analysis loses power (Field, 2012). Fortunately, SPSS statistical software includes Maulchy's test statistic which, if the assumption of sphericity is violated, provides alternative estimates for sphericity.

The Maulchy's test statistic in the repeated measures ANOVA indicated that the assumption of sphericity had been violated, χ^2 (27) = 61.33, $p = 0.00$, therefore degrees of freedom were corrected using Huynh-Feldt estimates of sphericity ($\varepsilon = .80$). According to Field (2013), when ε is $< .75$ then Huynh-Feldt is the most appropriate correction to use as opposed to Greenhouse-Geisser, which would be used when $\varepsilon > 0.75$. When correcting the degrees with freedom, the analysis showed that there was a difference between the means of each assessment within groups, $F(20.02) = 6.82$, $p < 0.01$. A summary of the within-subjects factors of the ANOVA can be seen in the table (Table 5) below. The table reports the adjusted values for the Huynh-Feldt*p* estimate of sphericity.

Table 5.

Repeated Measures Analysis of Variance.

Effect	*MS*	*df*	*F*	Huynh-Feldt*p*
Assessment Score	2904.33	6	26.66	<0.001
Assessment Score*Class	754.54	23.87	6.82	<0.001
Error	110.57	328.23		

Based on this result we can conclude that there are significant differences between assessment scores, as well as the average assessment score for each group as both *p* values are <0.001. This test also provides some insight into the measure of association as the Partial Eta Squared value (not shown in Table 4) = 0.323 which can be interpreted approximately 32.3% of the variance in the assessment score (outcome variable) is accounted for by the class of the learner. However, this *F* does not show which groups were significantly different from the comparison class, or whether the scores themselves differed at each measurement and therefore planned contrasts were used. The concept behind planned contrasts is to determine that the results for each assessment were statistically different at each measurement point when adjusting for class. In Table 6 this is illustrated. From this table it is clear that there was a significant difference between the average score at each measurement point when comparing Assessment Scores, apart from the contrast between measure 5 to 6, and 6 to 7. This means that the results of for assessment 5, on average, did not differ significantly from assessment 6, and the results for assessment 6, on average, did not differ significantly from assessment 7. This indicates that during that time period, where there should have been an increase in marks based on maturation and programme attendance, there was little observed increase across all participants.

Table 6.

Tests of Within-Subjects Contrasts for Assessment Scores

Source	Assessment Contrast	*MS*	*df*	*F*	*p*
Assessment Score	Measure 1 vs. Measure 2	6132.29	4	12.18	<0.001
	Measure 2 vs. Measure 3	860.43	4	6.82	<0.001
	Measure 3 vs. Measure 4	2215.83	4	16.63	<0.001
	Measure 4 vs. Measure 5	674.35	4	3.30	0.017
	Measure 5 vs. Measure 6	**304.80**	**4**	**1.55**	**0.20**
	Measure 6 vs. Measure 7	**447.52**	**4**	**2.50**	**0.053**
	Measure 7 vs. Measure 8	518.10	4	2.69	0.041

Note. The non-significant results are bolded, $p < 0.05$.

Even though there was no significant difference on average for each class in two of the measurements, this only highlights that on average the learners in those classes did not perform significantly better or worse than their previous score. It does not highlight the differences *between* the classes.

When looking at the between-subjects effects, the results showed that there was a significant difference between-subjects, $F(1,4) = 2188.25$, $p < 0.001$ and in order to find out which groups were different, and specifically which groups differed from the comparison group, two *post-hoc* tests were used, the Dunnett t and the Bonferroni *Post-Hoc* test. While the Bonferroni test is commonly used as a *post-hoc* test for ANOVAs, with the added advantage that it is widely acceptable and applicable to a multitude of contexts, it has the disadvantage that it is often lacking in power and weak (Newsom, 2006). Due to this, the Bonferroni *post-hoc* test was not the primary focus of the post-hoc analysis. However, some interesting results can be noted nonetheless, as seen in Table 7 below.

Table 7.

Performance Differences between groups using the Bonferroni Post-Hoc Test

(I) Class of Learner	(J) Class of Learner	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
School B, 2A	School B, 2B	13.1469*	4.23041	.030	.7742	25.5195
	School A, 2A	-13.5256*	4.05031	.015	-25.3716	-1.6797
	School A, 2C	16.7681*	4.23041	.002	4.3954	29.1407
	School A, 2B	12.8729*	4.13383	.029	.7827	24.9631
School B, 2B	School B, 2A	-13.1469*	4.23041	.030	-25.5195	-.7742
	School A, 2A	-26.6725*	4.23041	.000	-39.0452	-14.2998
	School A, 2C	3.6212	4.40315	1.000	-9.2567	16.4991
	School A, 2B	-.2740	4.31044	1.000	-12.8807	12.3328
School A, 2A	School B, 2A	13.5256*	4.05031	.015	1.6797	25.3716
	School B, 2B	26.6725*	4.23041	.000	14.2998	39.0452
	School A, 2C	30.2937*	4.23041	.000	17.9210	42.6664
	School A, 2B	26.3985*	4.13383	.000	14.3083	38.4887
School A, 2C	School B, 2A	-16.7681*	4.23041	.002	-29.1407	-4.3954
	School B, 2B	-3.6212	4.40315	1.000	-16.4991	9.2567
	School A, 2A	-30.2937*	4.23041	.000	-42.6664	-17.9210
	School A, 2B	-3.8952	4.31044	1.000	-16.5020	8.7115
School A, 2B	School B, 2A	-12.8729*	4.13383	.029	-24.9631	-.7827
	School B, 2B	.2740	4.31044	1.000	-12.3328	12.8807
	School A, 2A	-26.3985*	4.13383	.000	-38.4887	-14.3083
	School A, 2C	3.8952	4.31044	1.000	-8.7115	16.5020

*The comparison class is indicated by the blue font.

The Bonferroni test indicated that the comparison group was only statistically different to School A and B 2A ($p < 0.001$ and $p = 0.03$ respectively). Not only were these classes significantly different from the comparison class, but they also outperformed the comparison class on average by 26.67 and 13.15 points respectively.

As Newsom states (2006), the Bonferroni test is a useful test, but for repeated measures ANOVA there is a potential lack of power and therefore a threat of a Type II error. To minimize this threat, a second *post-hoc* test was also conducted. The Dunnett test is especially used when comparing multiple groups to one comparison group, and therefore is useful to validate the results of the Bonferroni. The results of the test can be found in Table 8 below.

Table 8

Performance Differences between comparison class and treatment classes

Measure	Class of Learner (I)	Comparison Class (J)	Mean Difference (I – J)	Sig.
Dunnett t (>control)	School A, 2A	School A, 2B	26.59	0.00
	School A, 2C	School A, 2B	-6.54	0.99
	School B, 2A	School A, 2B	12.02	0.016
	School B, 2B	School A, 2B	-1.69	0.90

This table confirms the results in Table 7 and highlights that the comparison class was only significantly different from two of the four treatment groups and it also outperformed the other two classes on average by 6.54 and 1.69 points in final year mark. This provided statistical power to the Figures presented earlier in the chapter (Figures 3 – 5).

These findings provided insight into how the classes were progressing in terms of mathematical ability, with two of the classes (School A, 2A and School B, 2A) significantly outperforming the comparison group and the other two classes (School A, 2C and School B, 2B) being outperformed by the comparison group. As a factor, these tables also show that the class of the learner is making a significant contribution to outcome variable of mathematical improvement, however when each class is considered as a part of the model, then only two of the classes are making a significant contribution to the model (School A, 2A and School B, 2A). While the above results confirmed this, a multiple regression model using the class of the learner as a dummy variable to turn the groups into usable dichotomous independent variables. Since multiple regression requires one dependent variable, the final year mark was used as a dependent variable was used for this analysis. The final year mark for each learner was made up of average score across the eight assessments for the year.

Table 9.

The effect of class on the final year mark of the learner.

R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					
				R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
.754[a]	.568	.536	11.03	.568	18.07	4	55	.000	1.24

The results of this analysis showed that the Class explained 56.8% of the total variation in final year for assessments, and this variation was statistically significant, $F(4,55) = 18.07$, $p < 0.001$. In essence, this meant that class was a better explanation for the outcome variable, final year mark, than no model at all. However, the regression model also identified which classes were making a significant contribution to the model, and these results can be seen in Table 10 below.

Table 10.

The contribution of explained variance by each class.

Model	Unstandardised Coefficients		Standardised Coefficients	t	*p*
	B	SE	*Beta*		
Constant (comparison group)	41.42	3.18		13.01	<0.001
School A, 2A	26.97	4.41	0.69	6.11	<0.001
School A, 2C	-6.14	4.60	-0.15	-1.33	0.19
School B, 2A	12.20	4.41	0.31	2.76	0.008
School 2, 2B	-1.32	4.60	-0.03	-0.29	0.77

These results highlighted that only two of the classes made a significant contribution to the model, namely the two A classes from each school. In summary, while class was a significant contributor to the final year mark of the learner, it appeared as though this was only true for learners that were part of the two A classes.

2. **Is there evidence that programme dosage is a significant contributor to the improvement of foundational numerical skills of the participants?**

The following figure (Figure 6) shows each of the 48 treatment class participants' number of sessions attended throughout the year The programme included 36 sessions in total. There was no enforcement of attendance, or required number of sessions attended for the programme.

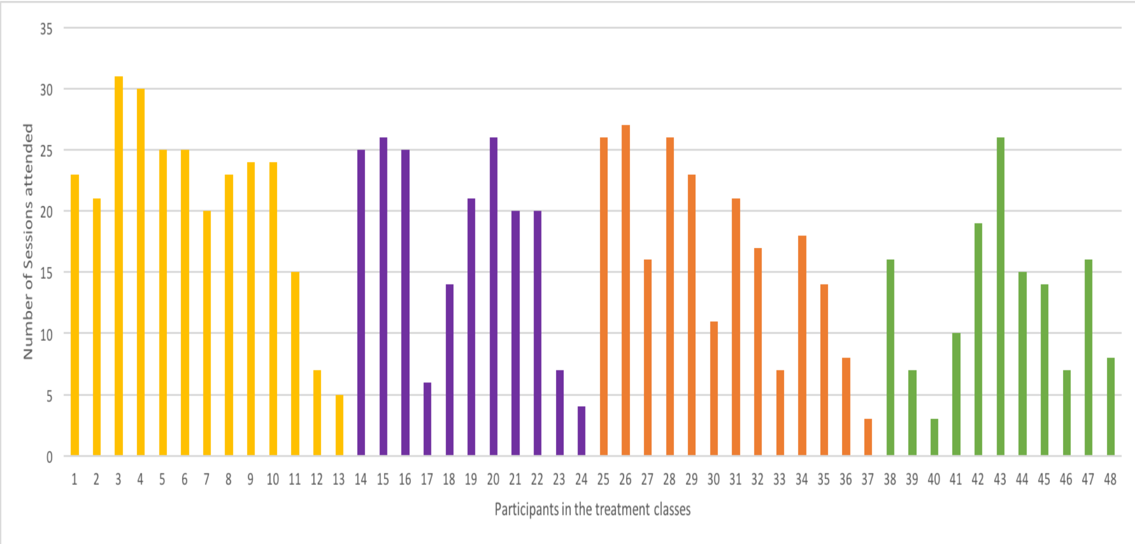

Figure 6. Treatment class participants' number of sessions attended.

Note. The x-axis indicates which class the learners belong to, and can be interpreted as follows:

- 1-13: Class 2A, School B
- 14-24: Class 2B, School B
- 25-37: Class 2A, School A
- 38-48: Class 2C, School A

At an initial glance this figure indicates that there was a large range in programme participation (31 - 3). In order to interpret this further, a comparison of means was done between classes and the number of sessions attended. The results can be seen in Table 11 below:

Table 11.

Number of Sessions Attended per treatment class

Class	Mean	N	Std. Deviation
Class 2A, School B	21.00	13	7.79
Class 2B, School B	17.64	11	8.48
Class 2A, School A	16.69	13	7.84
Class 2C, School A	12.82	11	6.59
Overall	17.19	48	8.02

The total number of planned sessions was 36, and the average learner (*n = 48*) attended 17.19 sessions, which is 47.78%. If the average learner is attending less than 50% of the programme, this is a cause for concern which will be addressed in the Discussion section below.

This comparison of means identifies a concern, that the class (2A, School A) that performed, on average, the best did not attend more than the overall average attendance. School 2A attended, on average, 16.69 sessions, which is 46.36% of all sessions. For further insight, the following figure displays each learner's final year mark against the number of sessions they attended:

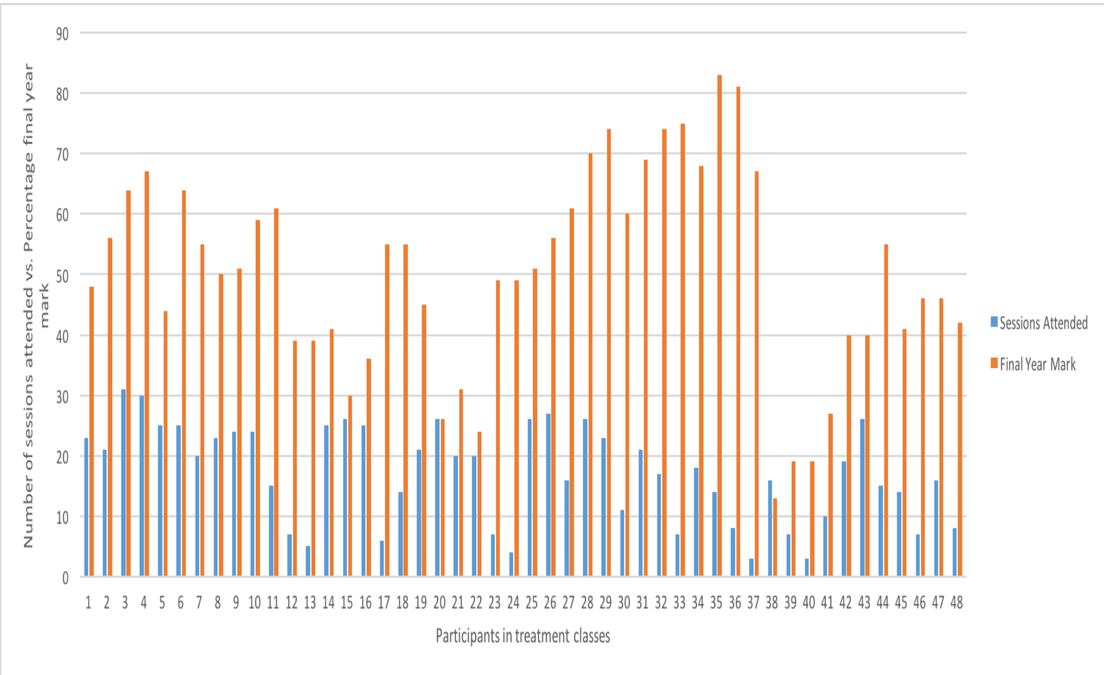

Figure 7. The comparison of final year mark again number of sessions attended.

While this does not provide a direct answer to the second evaluation question, it does provide a useful foundation to conduct further analyses. Interestingly, some of the best performing learners (36 and 37, from Figure 7), attended less than 10 sessions (less than 27.78%). In order to provide more nuanced results for the second evaluation question, a linear regression model was used with the final year mark for the learners as the outcome variable and the number of sessions attended as a predictor variable. The results can be seen in Table 12.

Table 12.
The effect of number of sessions attended on overall year mark for learners participating in the programme.

Model	*R*	*R Square*	*Adjusted R Square*	Std. Error of Estimate
1	0.210	0.044	0.027	15.97

The model highlighted that the number of sessions attended accounted for 4.4% of the variance in final year marks for learners that attended the programme, and it was not found to be a significant factor in the overall outcome, $F(1,58) = 2.67$, $p = 0.11$. It appears that attendance was not a good predictor of final year mark.

All these results will be discussed further in the Discussion Section below.

Discussion

The results will be discussed below, under the relevant evaluation question. Included in this discussion are the limitations of the evaluation, as well as a number of recommendations presented to help improve future iterations of the Numeracy Programme.

1. **Are the foundational numerical skills of the participants of the Numeracy Programme better after the programme than before, and better than those of learners who did not take part in the programme?**

The findings of this evaluation ultimately serve the purpose of determining whether or not the Numeracy programme was effective in reaching the short-term outcome of strengthening the foundational numeracy skills of the participants. The lack of consistent results between groups indicates that the programme did not inherently benefit the participants compared to those that did not receive treatment. This is not to say that there was no effect, as site differences were evident. The numeracy programme was effective in improving the Mathematical performance of learners in two of the four intervention groups. In these two classes, 2A in School A and B, the scores on the assessments were significantly different from the comparison group that did not participate in the programme. This is particularly interesting for Class 2A in School A, because the average score on the initial assessment in Term 1 was virtually the same as the comparison group (57.85 vs. 58.00), but by the end of the year the average score was more than 25% higher than the comparison group (78.74 vs. 47.83). All of the analyses indicated that this class was significantly different from the comparison class, and this seems to suggest that the numeracy programme was playing a role in the progression of these students. The concerning factor of this evaluation is that there was almost no difference between the other two classes, Class 2C in School A, and Class 2B in School B, and the comparison group. In fact, these intervention classes ended up performing worse on average than the comparison group by end-year. Needless to say that the results must be interpreted with caution, as there appears to be additional factors contributing to the outcome variable. If the programme brought about change, it would have been expected that there would have been significant differences between all intervention classes and the comparison class, and because this is not the case it is a challenge to make an argument that the differences that are presented are as a result of the programme alone. The mixed findings clearly show that despite the success of two classes, these participants appear to be not the target population of the programme.

Overcoverage

One of the core issues of this intervention is what is referred to by Rossi, Lipsey, and Freeman (2004) as overcoverage. Overcoverage can be understood as the number of participants that do not require the treatment of the programme compared with the total number of participants in the programme (Rossi, Lipsey, & Freeman, 2004). In this case, if one considers how different the results of School A, 2A and School B, 2A were then 26 of the participants in the programme were not actually in need (13 from each class). This indicates that 54.17% of the treatment participants should not have been receiving the treatment. The nature of overcoverage in a programme such as the Numeracy Programme is that it introduces extra cost, bias and increases the risk of error. Rossi, Lipsey, and Freeman (2004, p. 187) suggest three sources of information that can be used to determine if the programme is serving the target population:

- Programme Records
- Survey of programme participants
- Community surveys

The most appropriate source of information that would reduce overcoverage for the Numeracy Programme would fall under surveying the participants. Accurate testing or assessment to determine whether the programme is serving the intended participants can be considered a function of programme records and surveying programme participants. One recommendation would be for two standardised measures, one for selection of participants and one for mid-intervention to determine if the participants has met the required standard and can exit the programme. This evaluation was limited by the selection assessments used by the schools, as there was no way of knowing if the assessments were valid or reliable measures. The ideal for LMF in the future is to have the programme begin half way through the Grade 1 year and continue till the end of the second term of the Grade 2 year. In this way, there will be an exit for those learners who have achieved a sufficient level of improvement. An additional issue is to establish, prior to the beginning of the programme, whether classes are streamed according to cognitive ability. Even a reliable and valid selection assessment will not be able to provide a comparable baseline between high cognitive ability classes and low cognitive ability classes.

Response to Intervention

The Numeracy programme appears to operate under a response to intervention (RtI) model, a tiered model of instruction that includes three primary tiers, Tier 1 being the core-classroom experience, Tier 2 being additional instructional support in small groups, and Tier 3 being additional support in a one-on-one mode (Baker et al, 2014; Cirino et al, 2008). The idea is that the additional support provided by the Numeracy programme in small groups, in addition to the core classroom activities, will aid those who are struggling with mathematics (Baker et al, 2014). This model has become fairly standard in reading programmes, however there has been some difficulties in translating the RtI model for mathematics and, while there has been measured effectiveness, there has often been difficulty relating this measured effectiveness to one component of the intervention (Baker et al., 2014; Byrant et al., 2011). The difficulty that the Numeracy Programme faces is that this type of model is founded on high-quality classroom instruction (National Centre on Response to Intervention, 2010). This provides another explanation for the variation in results between classes being that the programme was not the only source of learning, and the regular classroom activities would have added to the knowledge base of the participants. Because each class did not have the same teacher, teacher competency is a potential confounding factor in the evaluation. Cirino et. al. (2008) has been cited earlier in this evaluation on their research relating to instructional design and how it plays an important role in eliminating and overcoming learning barriers in mathematics. If the classes that showed significant improvement were receiving higher quality teaching during school hours over and above the programme, this would explain the marked improvement in the participating learners. A limitation of the evaluation is the lack of a clear answer in this case.

Beckmann et al., (2009, pp. 11-12) provide teachers and school programmes with eight recommendations, in a checklist format, when considering a response to intervention model:

1. Screen all learners to identify those at risk for potential mathematical difficulties and provide the programme to the identified learners
2. Instructional materials for students receiving interventions should focus intensely on the concept of whole numbers and operations
3. Instruction during the programme should be explicit and systematic.

4. Programmes should include instruction on solving word problems that is based on common underlying structures
5. Programme materials and activities should include opportunities for students to work with visual representations of mathematical ideas.
6. Programme activities should devote 10 minutes in each session to building fluent retrieval of basic arithmetic facts.
7. Monitor the progress of learners receiving additional instruction (in the form of the programme) as well as the learners who are at risk
8. Include motivational strategies in Tier 2 and Tier 3 interventions.

This checklist, as adapted from Beckmann et al., contains information that could result in the improvement of the Numeracy programme in the following years. Based on the information and data collected, it is safe to say that the Numeracy Programme is meeting some, but not all, of those recommendations. Specifically, the first recommendation is not being met, and this is the most important one of them all. The successful administration of screening will resolve the issue of overcoverage and provide the programme with a much clearer picture of its effectiveness. Ultimately, the teacher's decision about a learner is subjective and, for the programme to improve, the decision needs to be based on a standard that can be applied to any school that LMF work in. As a result of this unreliable selection assessment, there may have been learners included in the programme that did not need to be there, and learners not included that should have been. In evaluating the findings of various studies, Beckmann et al. highlights that the most effective screening or selection measures for the primary school grades involve assessments of learners' knowledge of number comparison, operations, and strategic counting. These concepts form the basis of the South African curriculum, as outlined by CAPS (Department of Education, 2011). One such assessment that has been used and proposed by Beckmann, et al. is the *Numbers Knowledge Test*. This test is given to individual learners over 10-15 minutes and it assesses the learners' number concept knowledge and understanding as well as their understanding of number placement, relative size, and counting skills (Flojo, Gersten, & Jordan, 2005). It is suggested that assessments similar to this be consulted and examined when designing the selection assessment for the Numeracy Programme. Once this assessment has been designed, all learners who score below a certain percentage on the assessment will be included in the programme, rather than the weakest 12 learners from each class. This should also account for the differences in cognitive ability between classes.

Additionally, the eighth recommendation provides the learners with the positive reinforcement that plays an important developmental role in learners of this age (Beckmann et al., 2009). It is unclear whether this forms part of the Numeracy Programme, and therefore one suggestion is that the volunteers make an effort to reinforce or praise the learners when they succeed and allow them to keep track of their success via a chart or booklet (Beckmann et al., 2009).

Research from the National Council of Teachers of Mathematics (NCTM) suggest that a quality intervention should follow five recommendations (NCTM, 2007, p.2):

1. The small groups of learners should be no more than six learners in a group
2. Address skills that are necessary for the Grade of the learner and appropriate to classroom content
3. Content the sessions in a systematic and explicit manner
4. Require the learners to think aloud as they solve problems or use objects as visual representations
5. Have a balance of work and activities that are more complex as well as basic, as long as they are grade-specific problems.

From the resources and information provided it is clear that the Numeracy problem meets the first two requirements, however there was not enough evidence to conclude that the following three were met. This leads into the concept of instruction and volunteer proficiency, discussed in the section below.

Volunteer Proficiency

It is not only the quality of teaching that may have resulted in the mixed findings, but also the quality of instruction from the tutors might have been the issue. LMF volunteers are trained, but there is as yet no evidence of how well they have applied the training. It would be helpful to track volunteer competence after training and application of training by the volunteers.

The programme activities appear to be in line with current research around effective interventions, especially the use of numeracy-based board games and activities that provide an entertaining but educational aspect to the programme. If an intervention is not having the intended outcome it desires, but has a sound theoretical base, then it could point to an implementation error (Rossi, Lipsey, & Freeman, 2004).

Recommendations for first evaluation question

Overall, it needs to be highlighted that the findings of this evaluation indicate that the foundational numerical skills of the participants were not better after the programme than before, and they were not better than those that did not take part. Therefore, the following recommendations are provided based on findings and evaluations of similar programmes, and aim to improve the effectiveness of the Numeracy Problem:

1. The development of a standardised selection tool. An adequate baseline comparison would have indicated that the two A classes were already better off at start, and therefore should have been excluded.
2. The inclusion of positive reinforcement between volunteer and participant. Research has indicated that in small-groups at risk learners benefit from positive reinforcement and the means to track their progress. This provides a visual reference for the learners and motivates them to continue attending sessions.
3. Monitor and track the quality of volunteer instruction. Currently there is no evidence that supports the idea that the volunteers are effectively using the training they were provided. If such evidence can be provided for future evaluations, it would remove one alternative solution to the results.

2. Is there evidence that programme dosage is a significant contributor to the improvement of foundational numerical skills of the participants?

Based on the descriptive statistics, which indicated that the best performing class attended the programme less than the average learner, and the regression model, it is clear that programme dosage was not a significant contributor the improvement of foundational numerical skills. This further emphasises the issue of overcoverage, as there were more learners who did not need the programme than those that did. Returning to Rossi, Lipsey, and Freeman (2004), programme records are an important source of information regarding this. It is possible that the attendance records did not provide enough information around programme dosage, and therefore one suggestion is that more detailed information be kept on the topic rather than mere attendance. It is clear that unless an adequate assessment for selection is produced by LMF, the data they collect on programme dosage will remain untrustworthy and biased. In addition, programme attendance records fluctuated wildly, and it is not acceptable to expect that average attendance being below 50% will result in improvement, regardless of the quality of the programme. Programme dosage needs to be consistent and regular. The recommendations for this evaluation question echo those of the previous question, but in addition:

1. It is suggested that more detailed records on programme dosage be kept with regards to programme dosage and the participation of activities. It would be useful to see what each week consisted of in terms of activities, as well as the attendance of learners.
2. Attendance should be boosted so that the average attendance climbs well above 50%, and programme staff should motivate learners to attend more regularly. The previous recommendations mentioned the importance of motivation and positive reinforcement, and the same concept can be applied to the attendance of sessions. Learners that regularly attend should be rewarded, and those that do not should be receive follow up.

Limitations of the Evaluation

The evaluation was hampered by a number of limitations, such as the lack of an adequate comparison at baseline. This lead to the inclusion of learners that should not have been on the programme.

Secondly, there was incomplete knowledge around the academic characteristics of the classes. There seemed to be evidence supporting that the two A classes from each school were selected by academic merit, which would have provided a more nuanced insight into the effect of the programme. However, despite similar thoughts from the programme coordinator, one school (School A) stated that the classes were made up of differing academic ability, and the other school (School B) failed to respond to the query. This inconclusive evidence limited the interpretation of the results.

Finally, because the effect of normal classroom learning was not measured or accounted for in the research design, maturation became a serious threat to internal validity.

Contributions to knowledge

This evaluation has contributed to the planning and improvement of a pilot programme that was conducted for the first time in 2016. Even though the findings did not find a significant programme effect, the programme staff can implement standardised selection tools and thus create comparable intervention and control groups. Only then can one address the question of whether the programme caused improvement in foundational numeracy.

Conclusion

Overall, the evaluation found that the LifeMatters Numeracy Programme was not effective in improving the foundational numerical skills of its participants and there was no significant programme effect. However, part of the purpose of this evaluation was to aid programme improvement and identify focus areas for future iterations. In order for LMF to reach these outcomes there needs to be clarity around the academic level of the participants and how they were selected. An intervention such as the numeracy intervention is built on a strong selection process. For the next iteration of the programme it is vital that the correct students are selected, for both the treatment groups and the comparison group. The development of a relevant and appropriate numeracy assessment that is used to select participants should be a priority, as the subjective decision-making of teachers cannot be relied on.

The evaluation provided recommendations based on the response to intervention model, and various reports and reviews of that model. Tier 2, or supplementary interventions, have been shown to be effective in increasing the number concept, operations, and relationships skill, however this relies on a high-quality instructional design for the intervention that supplements the high-quality education received in the classroom. As the schools that LMF work with are under-resourced, it is suggested that further information on teacher competency, along with a valid and reliable selection measure, will ensure that the learners that need to be in the programme are included rather than those who do not. Once the issue of overcoverage is attended to, more accurate findings on programme effectiveness should be found, as long as the monitoring systems are expanded and improved at the same time.

Appendix

UNIVERSITY OF CAPE TOWN

School of Management Studies
University of Cape Town. Private Bag.
Rondebosch 7701
Telephone: +27 21 650-5218
Fax: +27 21 689-7570
8 February 2016

TO WHOM IT MAY CONCERN

Thank you very much for your willingness to enable one of our Master's students to work with a programme from your organization. I appreciate your contribution to the education of our students.

The student will need programme information from you and we request that you or a designated person meet with them regularly to provide access to this information. Your cooperation in this regard will ensure that the student meets deadlines and provides you with a high quality evaluation. In order to keep track of the student's interactions with your organization, we request that you copy the supervisor on all correspondence to the student.

Please note that our students are required to work within the ethical framework of the Faculty of Commerce when collecting information from programme documents or programme recipients. This framework deals with anonymity of data sources, sensitivity when requesting information from people and responsible reporting of results.

We also undertake and ensure you that the student will display professional behaviour at all times while working in your organization or on your programme. At the end of the process, you will receive a useful report which will enable you to make informed decisions regarding your programme.

In order to comply with the rules of the Faculty of Commerce, we request you to sign below to indicate that the student will have access to programme data and records and where applicable, to programme recipients.

Thank you very much.

Yours sincerely

J. Louw-Potgieter

PROF J LOUW-POTGIETER
CONVENER: MPHIL PROGRAMME EVALUATION

AGREEMENT TO ACCESS PROGRAMME RECORDS AND/OR RECIPIENTS:

[signature]	The LifeMatters Foundation	16/02/2016
AUTHORISED PERSON	ORGANISATION	DATE

References

Adelstein, S., Gormley, W., & Hill, C., (2012). Do the short-term effects of a strong preschool program persist. *Center for Research on Children in the United States, Working Paper, 18*, 17.

Adler, J., & Setati, M. (2000). Between languages and discourses: Language practices in primary multilingual mathematics classrooms in south africa. *Educational Studies in Mathematics, 43*(3), 243-269.

Aguilar, M., Marchena, E., Menacho, I., Navarro, J. I., Ruiz, G., & Van Luit, J. E. (2012). Longitudinal study of low and high achievers in early mathematics. *British Journal of Educational Psychology, 82*(1), 28-41.

Babbie, E. R. (2013). *The basics of social research.* Belmont: Cengage Learning.

Bailey, D. H., Clements, D. H., Domina, T., Jenkins, J. M., Nguyen, T., Sarama, J., . . . Spitler, M. E. (2015). Fadeout in an early mathematics intervention: Same old schools or underlying skills?. *Society for Research on Educational Effectiveness,*

Baker, S., Cary, M.S., Clarke, B., Doabler, C.T., Fien, H., Kosty, D.,& Smolkowski, K. (2014). Preliminary evaluation of a tier 2 mathematics intervention for first-grade students: Using a theory of change to guide formative evaluation activities. *School Psychology Review, 43*(2), 160.

Barnett, W. S. (1995). Long-term effects of early childhood programs on cognitive and school outcomes. *The Future of Children,* , 25-50.

Barnett, W. S. (2011). Effectiveness of early educational intervention. *Science (New York, N.Y.), 333*(6045), 975-978. doi:10.1126/science.1204534 [doi]

Baroody, A. J., Burchinal, M., Carver, S. M., Frye, D.,Jordan, N. C., & McDowell, J. (2013). *Teaching math to young children: A practice guide.* (No. NCEE 2014-4005). Washington, DC: National Center for Education Evaluation and Regional

Assistance (NCEE), Institute of Education Sciences, U.S. Department of Education.

Beckmann, S., Clarke, B., Foegen, A., Gersten, R., Marsh, L., Star, J. R., & Witzel, B. (2009). Assisting students struggling with mathematics: Response to intervention (RtI) for elementary and middle schools. NCEE 2009-4060. *What Works Clearinghouse.*

Bryant, J. D., Compton, D.L., Fuchs, D., Fuchs, L.S., Hamlett, C. L., & Paulsen, K. (2005). The prevention, identification, and cognitive determinants of math difficulty. *Journal of Educational Psychology, 97*(3), 493.

Bryant, D. P., Bryant, B. R., Chavez, M. M., Gersten, R., & Scammacca, N., (2008a). Mathematics intervention for first-and second-grade students with mathematics difficulties: The effects of tier 2 intervention delivered as booster lessons. *Remedial and Special Education, 29*(1), 20-32.

Bryant, D. P., Bryant, B. R., Funk, C., Gersten, R. M., Pool, C., Scammacca, N. N., & Winter, A. (2008b). The effects of tier 2 intervention on the mathematics performance of first-grade students who are at risk for mathematics difficulties. *Learning Disability Quarterly*, *31*(2), 47-63.

Cirino, P. T., Fletcher, J.M., Fuchs, D., Fuchs, L.S., Powell, S.R., & Seethaler, P.M. (2008). Intensive intervention for students with mathematics disabilities: Seven principles of effective practice. *Learning Disability Quarterly, 31*(2), 79-92.

Claessens, A., Chen, M., Davis-Kean, P. E., Duckworth, K., Duncan, G. J., Engel, M., . . .Siegler, R. S. (2012). Early predictors of high school mathematics achievement. *Psychological Science, 23*(7), 691-697.

Clements, D. H., & Sarama, J. (2011). Early childhood mathematics intervention. *Science (New York, N.Y.), 333*(6045), 968-970. doi:10.1126/science.1204537 [doi]

Davis-Kean, P. E., Duncan, G.J., Siegler, R.S., & Watts, T.W. (2014). What's past is prologue: Relations between early mathematics knowledge and high school achievement. *Educational Researcher*, *43*(7), 352-360.

Department of Education. (2011). *Curriculum and assessment policy statement grades 1-3.* Unpublished manuscript.

Department of Basic Education. (2012). *Annual Report: 2012/13*. Pretoria: South Africa. Retrieved from: http://www.education.gov.za/Portals/0/Documents/Reports/DBE%20Annual%20Report%202012-13.pdf?ver=2015-01-30-082145-573

Department of Basic Education. (2014). *The annual national assessment of 2014: Diagnostic report foundation phase mathematics and home language*. Pretoria: Republic of South Africa.

Dowker, A., Heine, A., Kadosh, R. C., Kaufmann, L., & Kucian, K. (2013). Interventions for improving numerical abilities: Present and future. *Trends in Neuroscience and Education, 2*(2), 85-93.

Donaldson, S.I. (2007). *Program theory-driven evaluation science: Strategies and Applications.* Mahwah, NJ: Erlbaum.

Eksteen, L. J. (2014). *Mathematical Learning Difficulties in Grade 1: The Role and Interrelatedness of Cognitive Processing, Perceptual Skills and Numerical Abilities,*

Field, A. (2013). *Discovering statistics using IBM SPSS statistics.* Los Angeles: Sage.

Flojo, J., Gersten, R., & Jordan, N.(2005). School readiness and later achievement. *The Journal of Learning Difficulties,* 38(4), p. 302.

Fleisch, B. (2008). *Primary education in crisis: Why south african schoolchildren underachieve in reading and mathematics* Juta and Company Ltd.

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

Fuchs, L. S., Fuchs, D., & Hollenbeck, K. N. (2007). Extending responsiveness to intervention to mathematics at first and third grades. *Learning Disabilities Research & Practice, 22*(1), 13-24.

Hatry, H. P., Newcomer, K. E., & Wholey, J. S. (2015). *Handbook of practical program evaluation.* John Wiley & Sons.

Howie, S. J. (2003). Language and other background factors affecting secondary pupils' performance in mathematics in South Africa. *African Journal of Research in Mathematics, Science and Technology Education, 7*(1), 1-20.

Jordan, G. B., & McLaughlin, J.A. (2010). Using logic models. In J. S. Wholey, H. P. Hatry, & K. E. Newcomer (Eds.), *Hand Book of Practical Program Evaluation* (3rd ed.) (pp. 55–80). San Francisco, CA: Jossey-Bass.

LifeMatters Foundation. (2015). *LifeMatters Foundation Annual Report 2015.* Retrieved from LifeMatters Foundation Website: http://www.chicmamasdocare.org.za/uploads/5/4/0/1/54019901/lifematters_annual_report.pdf

LifeMatters Foundation. (2016). *LifeMatters Foundation Annual Report 2016.* Retrieved from LifeMatters Foundation Website: http://lifemattersfoundation.org/EN/wp-content/uploads/2012/10/2016-Annual-Report-LowRes.pdf

National Center on Response to Intervention. (2010). *Essential Components of RTI – A Closer Look at Response to Intervention.* US Department of Education. Retrieved from: https://usm.maine.edu/sites/default/files/smart/rtiessentialcomponents_051310.pdf

National Council of Teachers of Mathematics. (2007). *Mathematics reaching today: Improving Practice, improving student learning.* Reston, VA.

Newsom, J. (2006). *Post-hoc tests.* Unpublished manuscript.

Padgett, D. K., Royse, D., & Thyer, B. A. (2010). *Program evaluation: An introduction to an evidence-based approach* (5th ed.). Belmont: Cengage Learning.

Ramani, G. B., & Siegler, R. S. (2011a). Promoting broad and stable improvements in low-income children's numerical knowledge through playing number board games. *Child Development, 79*(2), 375-394.

Ramani, G.B., & Siegler, R.S. (2011b) Reducing the gap in numerical knowledge low- and middle-income preschoolers. *Journal of Applied Development Psychology*. *32,* 146-159.

Spaull, N. (2013). Poverty & privilege: Primary school inequality in south africa. *International Journal of Educational Development, 33*(5), 436-447.

van Wyk, A. (2014). English-medium education in a multilingual setting: A case in south africa. *International Review of Applied Linguistics in Language Teaching, 52*(2), 205-220.