BASIC CONCEPTS PROGRAMME

EVALUATION FINAL REPORT

SEPTEMBER, 2022

PREPARED BY

CHRISTOPHER ANDERSON, JESSICA HORLER & MATTHEW SNELLING

ALACRITY development

Basic Concepts unlimited
Building Thinking - Harnessing Potential

DG MT

TABLE OF CONTENTS

INTRODUCTION

This report presents the outcome evaluation of the Basic Concepts Programme (BCP). The evaluation began in 2020 and was completed in 2022, in the Northern Cape. The evaluation was commissioned by the DG Murray Trust to assess the extent to which Basic Concepts improves the school preparedness of Grade R learners attending schools that use BCP-trained teachers, who implement the BCP curriculum. The evaluation makes use of a treatment group and comparison group.

The primary evaluation question that the evaluation seeks to answer is as follows: *Does the Basic Concepts Programme improve the school preparedness of programme learners at the end of Grade R, and if so, by how much?*

A baseline assessment of learner development was conducted in the beginning of 2020 using the Early Learning Outcomes Measure (ELOM). The evaluation initially intended to test the same groups of learners at the end of 2022 using the Test of Basic Concepts Knowledge (TBCK). However, due to the COVID-19 pandemic and resultant lockdowns, this was not possible. Endline fieldwork resumed in 2021, where TBCK assessments were completed on a different cohort of learners from the same BCP schools and the same comparison schools. As such, ELOM and TBCK results cannot be directly compared for individual children. Rather, the baseline study is used to establish equivalent performance between treatment school children and comparison school children, thus, allowing us to identify differences in the TBCK performance between these two groups at endline.

ABOUT THE BASIC CONCEPTS PROGRAMME

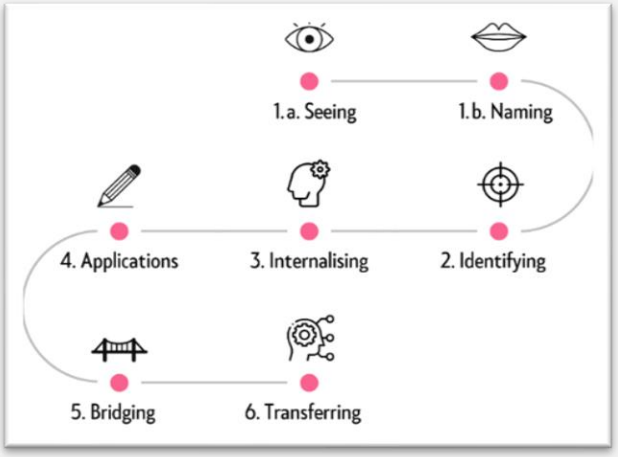
The BCP is an evidence-based metacognitive programme for young children facing language, information processing, and socio-emotional barriers towards learning. It is designed to be a short-term, intensive, small-group and semi-structured intervention (Benjamin, 2005). The programme is based on contemporary cognitive educational theory, drawing on the works of developmental theorists Jean Piaget, Lev Vygotsky, and Reuven Feuerstein (Benjamin, 2005).

The programme was developed to address the specific cognitive and developmental needs of young children as they transition from the concrete thinking of pre-school to the cognitive capabilities required for formal learning (reading, writing, spelling, and mathematics).

The BCP takes a scaffolded approach in which learners are introduced to hierarchically constructed conceptual systems that provide a foundation for a series of cognitive capabilities (Benjamin, 2005).

The BCP content is mediated using six teaching steps, followed by teachers who have been trained to deliver the programme using a mediational teaching approach (see Figure 1).

Figure 1. BCP Teaching Model.



Teachers are encouraged to develop a learning relationship with children, with the goal of assisting them to structure their thinking and develop the cognitive tools needed for learning.

The programme routinely makes use of the Test of Basic Concepts Knowledge (TBCK), a developmental assessment designed by Dr Benjamin, which assesses learner knowledge of six conceptual domains (colour, shape, size, position, number, and letter). This tool is used to measure child performance at the start and end of the year.

EVALUATION METHOD

A quasi-experimental evaluation design using multi-stage cluster random sampling was used. A comparison group was sampled from schools who were not yet receiving the intervention, but who were candidates from the same communities for future intervention. All intervention schools had been participating in the programme for at least one complete year, and teachers within intervention schools had completed BCP training. The best evidence available to the programme suggests that the schools in these communities serve the same socio-economic strata of children, and thus, all schools used in this evaluation (both groups) should serve children from similar backgrounds with similar developmental trajectories.

The evaluation employed a baseline assessment and endline assessment to measure performance of the treatment and comparison groups across time. At baseline, Grade R groups were measured using the Early Learning Outcomes Measure (ELOM), and at endline, the groups were measured using the BCP's TBCK for Grade 1 learners. The baseline assessment serves as a measure of equivalence between the two groups, as neither group of children had yet been exposed to any intervention teaching. Only children from intervention schools would then receive attention from BCP-trained teachers, and exposure to the BCP curriculum.

Changes to the Evaluation Method

The initial evaluation proposal intended to compare baseline ELOM scores with endline TBCK scores of the same group of children. However, this was not possible. Baseline fieldwork was completed in February 2020, right before the onset of the COVID-19 pandemic and associated lockdown periods. Consequently, the endline fieldwork that was planned for the end of the year had to be cancelled. The BCP was not implemented as intended in 2020, and fieldwork was not possible. Thus, the design was amended to allow the measurement of different cohorts of children, but from the same groups of schools in the same communities. In addition, we added an analysis of cross-sectional data over time for intervention and control schools using historical TBCK data. This means that this evaluation now includes elements of time-series and cross-sectional design. This requires that all phenomena affecting one group of schools

over time (including features of the pandemic) are expected to affect all schools equally. As all treatment and comparison schools were no-fee schools drawn from the same districts, children beginning at these schools would not be expected to perform differently year on year. Final endline data collection took place in February and March 2022.

Measurement Tools

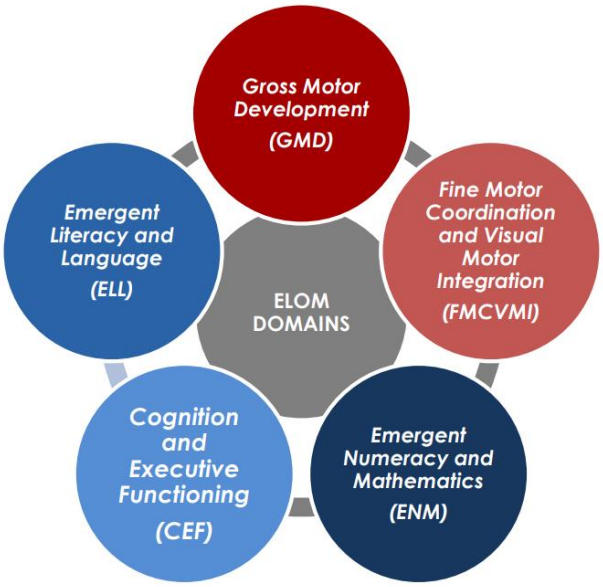
Early Learning Outcomes Measure

Baseline data collection was conducted using the Early Learning Outcomes Measure (ELOM). The ELOM is a population-level instrument designed to measure the developmental status of children aged 50 to 69 months. Children are individually assessed by trained assessors in their home language in a session lasting about 45 minutes. Scores are captured on a tablet and uploaded to a server for cleaning and analysis.

The ELOM consists of 23 direct assessment items clustered in five domains (see Figure 2). During standardisation, psychometry based on Item Response Theory was conducted on a sample of 1331 children from five language groups and representative of five socio-economic strata[1]. The ELOM is a reliable, age valid tool that provides a fair assessment of children from across ethnolinguistic groups. Children's expected performance is specified in Early Learning Development Standards (ELDS)[2]. These are available for ELOM total scores and for each domain. Full details may be found at: http://elom.org.za. For this study, we permit the expansion of the ELOM age range up to 71 months of age.

[1] Snelling, M., Dawes, A., Biersteker, L., Girdwood, E., & Tredoux, C. (2019). The development of a South African Early Learning Outcomes Measure: A South African instrument for measuring early learning program outcomes. Child: Care, Health and Development, 45, 257–270.
[2] Dawes, A., Biersteker, L., Girdwood, E., Snelling, M.J.T.L., Tredoux, C.G. et al. (2020). Early Learning Outcomes Measure (ELOM) technical manual. The Innovation Edge. http://ELOM.org.za/wp-content/uploads/2020/06/ELOM-Technical-Manual_2020.pdf

Figure 2. Developmental Domains Assessed by the ELOM.



Test of Basic Concepts Knowledge

Endline data collection was conducted using the BCP's TBCK. The TBCK was developed by BCP's founder, Louis Benjamin, to determine the preparedness of learners at the end of Grade-R and at the beginning of Grade 1[3]. The instrument can also be used to identify learning barriers in Grades 2 and 3 learners. The TBCK assesses knowledge in six conceptual domains: colour; shape; size; position; number; and letter. The TBCK places learners into four distinct performance bands according to their assessment scores: very weak (scores between 0 – 10); weak (scores between 11 – 17); average (scores between 18 – 20); and strong (scores between 21 – 24). Learners who score in the weak and very weak categories are deemed at risk of lacking the preparedness required to properly benefit from the next phase of schooling. The TBCK can be used to screen children for the programme, and also as a post-test to track improvements (as was done in this evaluation).

[3] Benjamin, L. (2017). *Test of Basic Concepts Knowledge (TBCK-R)* [seminar presentation].

Sampling

Baseline Sampling (ELOM)

The evaluation took place in the John Taolo Gaetsewe (JTG) and the Pixley Ka Seme (PKS) districts of the Northern Cape. Multi-cluster sampling took place, as follows: (1) All Afrikaans schools in PKS (13) were selected, and all Setswana schools in JTG (15) were selected. (2) Schools were randomly selected in each district by assigning them a random number and then ranking them from lowest to highest number. The top 10 schools from these lists were chosen for assessments. (3) Once the assessors arrived at the school, they obtained class lists from each BCP-trained teacher (treatment group), or comparison schoolteacher. Children were then randomly selected; 8 children were selected from each intervention school, and 4 children were selected from each comparison school. Children outside of the ELOM range were not assessed, however, the age was extended to 71 months. The total target number of assessments was 240, and the total number of assessments conducted during fieldwork was 244 (See Table 1).

Upon completion of fieldwork, all assessments were downloaded from the server and checked according to the criteria listed in Table 2. A total of 4 assessments were removed.

Table 1 displays the total number of assessments that were conducted during baseline fieldwork, in relation to the sampling targets. These assessments were analysed and reported on in the BCP baseline report. Upon completion of fieldwork, all assessments were downloaded from the server and checked according to the criteria listed in Table 2. A total of 4 assessments were removed.

Table 1. Target and Realised Sample - Baseline.

Intervention Group	Districts	Schools per District Visited at Baseline	Schools per District Visited at Endline	Number of Baseline ELOM Assessments
Treatment	PKS	10	7	84
	JTG	10	8	80
Comparison	PKS	5	3	40
	JTG	8	5	40
TOTAL		**33**	**23**	**244**

Table 2. Criteria for Data Removal.

Criterion	Number of Cases Removed at Baseline
1. The child failed the WHO disability screening[4]	0
2. The child refused to participate after the assessment had already begun.	0
3. The assessment was judged invalid if the child had a Total ELOM score < 15 and a Task Orientation score = 0. [5]	2
4. The child was assessed but was not of appropriate age.	1
5. The child was assessed in a language that was not their home language.	0
6. Duplicate data was submitted or was a trial.	0
7. The assessment was compromised due to assessor error.	0
8. The child failed two or more domains and scored a 0 for Task Orientation.	1
Total cases removed	4

[4] The adapted WHO screening tool for use with the ELOM is explained in the Technical Manual.
[5] ELOM Task Orientation is explained in the Technical Manual.

CALCULATING STATISTICAL POWER

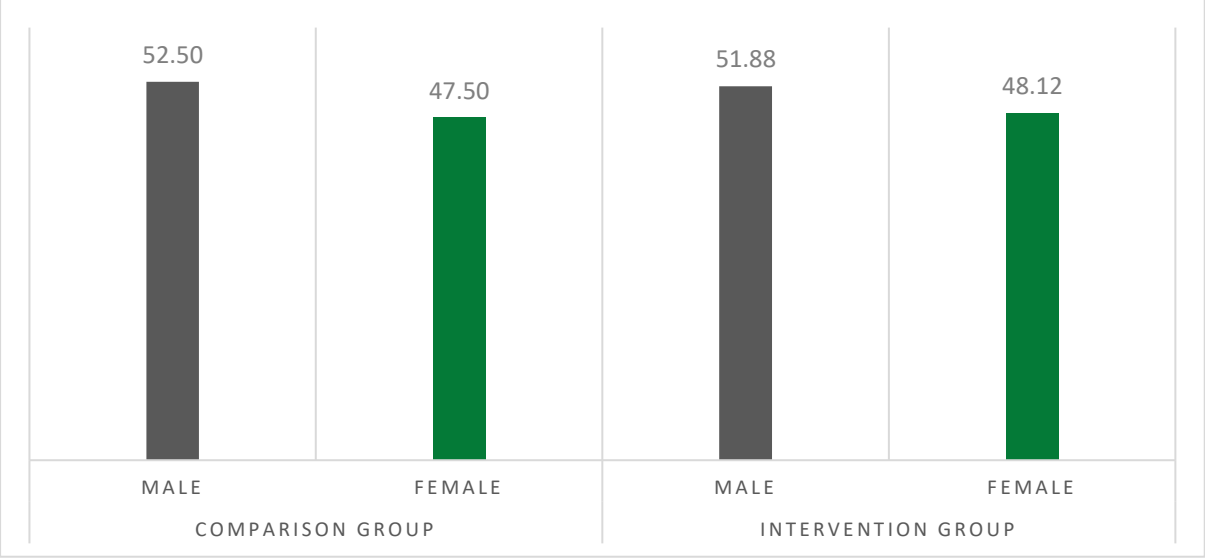
Using G*Power, the following statistical power was calculated for the target baseline sample of *n* = 240.

- Analysis: Analysis of Variance with fixed effects and interactions.
- Effect Size = 0.25
- Power = 0.80
- Alpha = 0.05
- Sample N = 196
- Adjustment for Attrition = 1.20
- Adjusted Sample N = 235 (240 for sampling purposes)

Baseline Sample Characteristics

The baseline sample was evenly split in terms of gender, with 52% of the sample being female, and 48% of the sample being male. These statistics are presented per intervention group in the table below, showing little difference between the Intervention Group schools and Comparison Group schools before being exposed to the programme activities.

Figure 3. Gender Split per Treatment Group.



The average age of the children in the sample was 65.2 months old (5.43 years), with the youngest child being 55.7 months old (4.64 years), and the oldest child being 70.9 months old (5.91 years). These statistics are presented per treatment group in the table below, showing little difference between the groups.

Table 3. Age Statistics per Intervention Group.

	Min Age	Max Age	Average Age	Standard Deviation
Comparison Group	55.9	70.8	65.4	3.3
Intervention Group	55.7	70.9	65	3.6
	55.7	**70.9**	**65.2**	**3.5**

Height-for-age z-scores (HfAZ) indicate whether a child is unusually short for their age. A z-score of -2.00 or less is an indication of possible stunting. At this time, the tool devised for generating these scores has only been expanded to a maximum age of 68 months. As such, the figures below refer to all children in the sample aged 68 months or below (*n* = 208).

The figures below serve as an indication of the proportion of potentially stunted children in the BCP intervention sample and comparison sample. We note a very similar proportion of children in the treatment and comparison sample are stunted.

Figure 4. Percentage of Children Stunted per Treatment Group (50 – 59 Month Group).

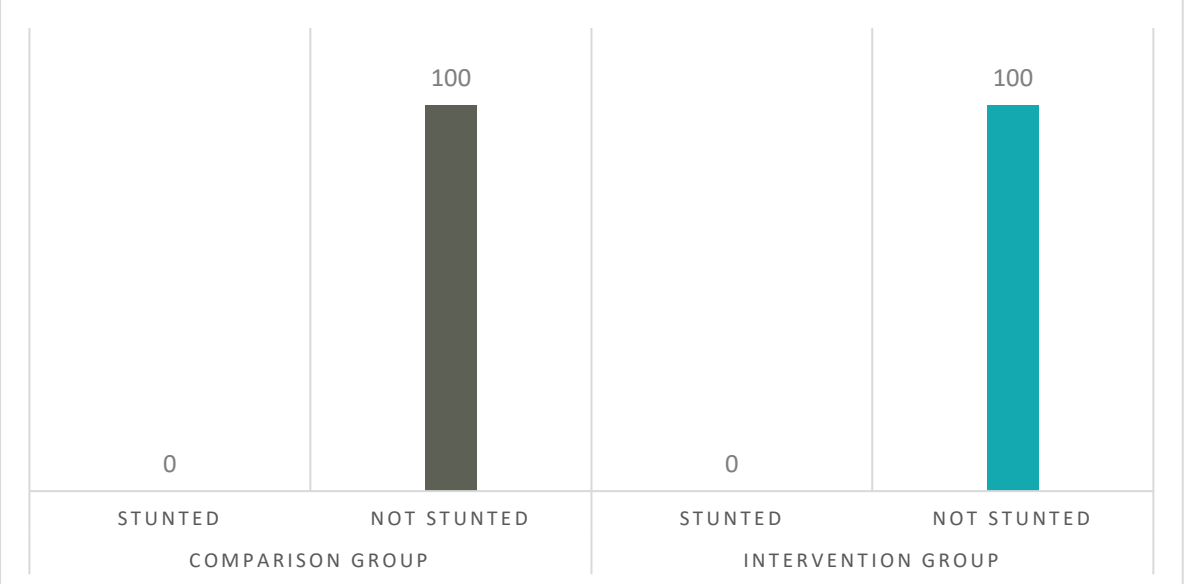
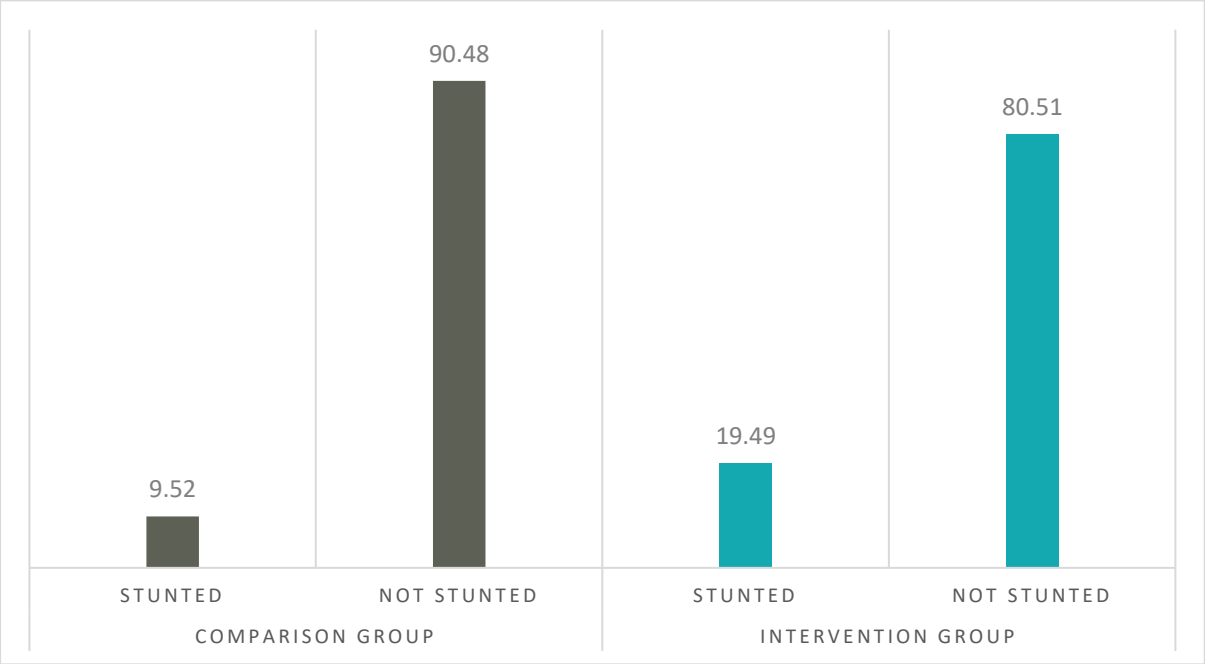
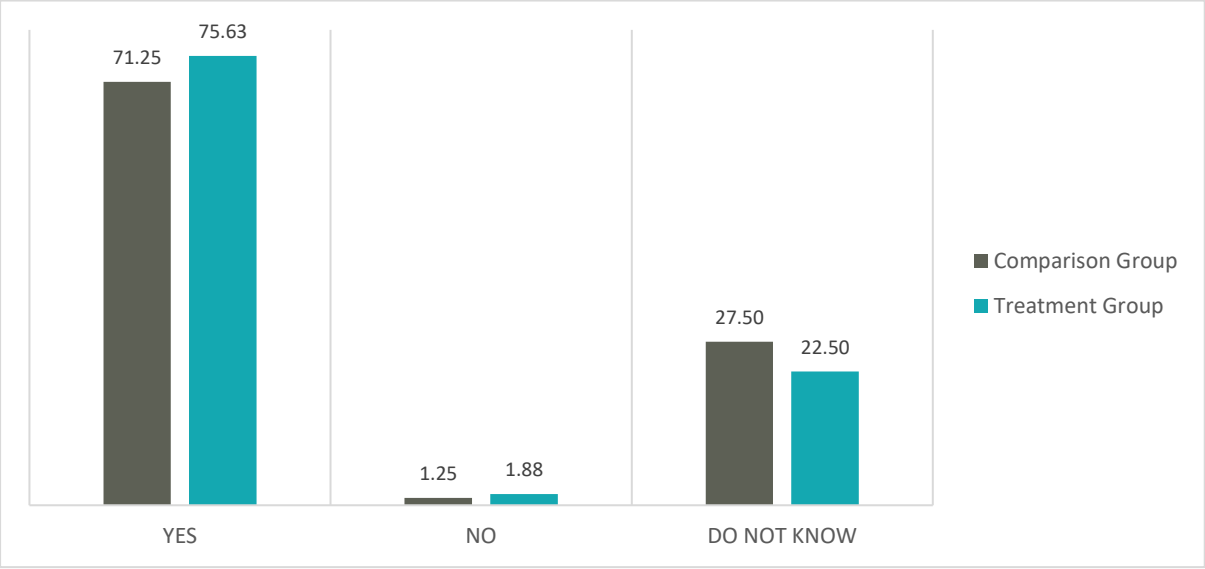


Figure 5. Percentage of Children Stunted per Treatment Group (60 - 68 Month Group).



The majority of the baseline evaluation sample (74%) receive the Child Support Grant. More children in the intervention group receive the grant (n = 121; 76%) compared to the comparison group (n = 57; 71%). However, given that the comparison sample is half the size, these two are proportionately the same.

Figure 6. Percentage of Children in Receipt of the Child Support Grant.



Endline Sampling (TBCK)

Endline fieldwork revisited the same schools as the baseline, with the exception of 13 schools that were not visited due to conditions surrounding fieldwork (six in PKS and seven in JTG). Further changes needed to be made to the school groupings (Intervention/Comparison) due to changes that took place at the schools during the evaluation period. These changes are as follows:

- A PKS comparison school was removed from the sample due to potential contamination.
- A JTG comparison school was reclassified as an intervention school at endline, as a Grade R teacher, who had been previously trained, started implementing the BCP programme during the evaluation period.
- A PKS intervention school was reclassified as a comparison school as the BCP-trained teacher left the school.

Based on these changes, the baseline ELOM dataset was amended and re-analysed for this report. The amended baseline and endline assessment numbers (after cleaning) are shown in Table 4.

As with the baseline ELOM fieldwork, the random selection of children also took place at the classroom-level and was conducted by the TBCK assessors.

All baseline descriptive statistics (Gender, Stunting, and Child Support Grant) were reconfirmed with the matched sample described below. The endline sample was almost evenly split in terms of gender, with 52% of the sample being female, and 48% of the sample being male. The average age of the learners in the sample was 77.2 months old (6.43 years), with the youngest learner being 6 years old, and the oldest learner being 8 years old. In the intervention group, 49% of learners were male and 51% female. Intervention group learners average age was 76.6 months old (6.38 years) with the youngest being 6 years old and the oldest being 7 years old. The comparison group had a higher percentage of female learners with 55% compared to 45% being male. The average of comparison learners was 78.72 months (6.56 years), with the youngest being 6 years old and the oldest being 8 years old.

The nature of the sample in these respects,

did not change meaningfully. Figures for

Endline Sample Characteristics are presented in Appendix A.

Table 4. Amended Baseline and Endline Samples Included in Final Analyses.

Intervention Group	Districts	Baseline Schools per District	Number of Baseline ELOM Assessments	Endline Schools per District	Number of Endline TBCK Assessments
Treatment	PKS	7	60	7	99
	JTG	6	51	6	55
Comparison	PKS	1	8	1	14
	JTG	5	28	5	50
TOTAL		**19**	**147**	**19**	**218**

Data Collection and Fieldwork

Baseline

Baseline ELOM data was collected in February 2020 by 4 trained ELOM assessors. Fieldwork was managed by Alacrity Development. Two assessors were based in PKS and spoke Afrikaans, and the other two were seTswana-speaking, based in JTG. A total of 33 schools were visited during fieldwork. Consent to participate in the evaluation was obtained from each school, and district officials in both PKS and JTG. Assessments took place in classrooms where the assessor could be alone with the child in an approximate 45-minute session.

Endline

Endline TBCK assessments were conducted from February to March 2022 by trained BCP assessors. Fieldwork was facilitated by Basic Concepts Foundation (BCF) supervised by Alacrity Development. Four assessors were based in JTG and three in PKS, again, assessing in Afrikaans and seTswana in the school's classrooms. Unlike the ELOM, the TBCK is administered as a group. Children are seated in rows one behind each other. Instructions are given from the front of the room using an enlarged script to demonstrate, while the second assessor ensures that all children are orientated and on-task.

Data Analysis

Baseline

ELOM assessments were submitted to a central server by the assessors every day during fieldwork. During this time, the assessments were checked daily for issues that might necessitate removing the assessment (and, therefore, reducing sample size), such as failing a disability screen or assessors assessing children outside of the ELOM age range.

Upon completion of fieldwork, all assessments were downloaded from the server and data cleaning commenced. A total of three assessments were removed: one was removed because the child scored less than fifteen points on the total ELOM score and failed the Task Orientation scale[6]; one was removed because the child failed two ELOM domains and failed the Task Orientation scale; and one was removed because they were not the right age for ELOM assessment.

After cleaning, a total of 147 ELOM assessments were included in descriptive analyses. Of this total, 36 assessments were from comparison schools and 108 assessments were from intervention schools. Descriptive statistics were performed. Average ELOM scores were compared to the ELOM standards to understand whether children in the baseline sample were *At Risk* of developmental delay, were *Falling Behind* what is expected of them at their age (the ELOM standard), or whether they were *Achieving the Standard,* and on track and needing no intervention.

Endline

TBCK assessments were handed over to the evaluation team at the end of March 2022. The dataset was cleaned using IBM SPSS and Microsoft Excel. Five cases were removed from a JTG intervention school due to poor school attendance.

A total of 218 TBCK assessments were included in the analyses (154 intervention learners and 64 comparison learners). Descriptive analyses were performed to determine learners'

[6] Task Orientation Scale may indicate inattention and poor engagement with the assessment.

performance in relation to the TBCK performance bands (*very weak, weak, average, strong*). Independent sample t-tests were then conducted using IBM SPSS to investigate differences in performance between intervention and comparison learners.

Timeseries

All endline TBCK data were provided by the Basic Concepts Foundation. This data referred to Intervention and Comparison schools from 2007 until 2022. There were areas where either an intervention school was not measured during a particular year, or where a comparison school was not measured. In order to analyze this data, we produced means per year and proportions of children who met the standards of the programme – that is scoring Average or Strong on the TBCK.

We then conducted independent sample t-tests for the entire range of children to determine whether there was a statistically significant difference between intervention and comparison schools over time, and produced Cohen's D statistics to indicate how different the groups were across the entire distribution.

GROUP EQUIVALENCE

A cornerstone of this evaluation is the equivalence between the intervention group and comparison group. For our conclusions to be valid, we must establish that, before intervention is delivered, intervention and control schools perform similarly. Baseline data collection established that, according to ELOM performance, intervention and control schools did indeed perform similarly, however, at endline, some schools were not able to be measured. Thus, it was necessary to look again at the baseline performance for those schools that could be matched to endline data and compare them to the original baseline results. This process of comparison is described in this section.

Baseline ELOM Performance

In this section, we present the average ELOM domain scores for the evaluation sample who are categorised as between the ages 60 to 71 months old (so as to correspond to the ELOM performance bands). It is important to note that these findings are descriptive only and do not represent the influence of other variables, such as teacher factors or characteristics of the district. Their purpose is to provide a simple overview of baseline performance on the ELOM.

The tables that follow use the ELOM convention for colour-coding the cells:

	Children are *At Risk*
	Children are *Falling Behind*
	Children are *Achieving the ELOM Standard*

For all comparisons, children in the 50 – 59 month groups were excluded, as these groups were very small, especially within the comparison groups. ELOM results are typically not analysed for groups with less than 15 children. These children cannot be combined with the 60 – 70 age group as their scores were markedly different. An independent samples *t*-test was performed to test this, showing that the younger children are indeed significantly different from the older group ($t = 3.8$; $p < .001$). Consequently, this group was removed.

THE ELOM STANDARDS

The expected ELOM performance standards are benchmarked at the standard score achieved by the top 40% of children (the 60th Percentile on the distribution). Thus, those who are At Risk are in the bottom 32% (below the 32nd percentile), and are well below the standard and need significant assistance to come up to the standard. Children who are Falling Behind are in the middle (between the 32nd and 59th percentile), performing better than those who are At Risk but not as well as those in the top 40%; with support they should be able to achieve the standard.

Children's overall ELOM scores do improve with age. This is to be expected as they learn and develop. However, their position within the ELOM performance bands does not change as a function of their increasing age alone, but rather, as a function of enhanced learning opportunities (such as participation in BCP).

For the unmatched baseline groups in the older age bracket, performance was similar between intervention and comparison. Both were *Achieving the Standard* on Gross Motor Development, and were *Falling Behind* on all other domains. The exception being Emergent Numeracy and Mathematics, where the intervention group was *Achieving the Standard*[7], scoring 0.41 points higher than the comparison group.

In practical terms, there is no difference between the comparison and intervention groups. This was confirmed by an independent samples *t*-test ($t = <1.5$; $p > .05$).

[7] In order to *Achieve the Standard* on Emergent Numeracy and Mathematics, the average score needs to be 10.24 points and above. Therefore, the intervention group only just qualify for this performance band.

Table 5. Average ELOM Scores for the 60 – 70 Month Age Group.

	Comparison Group (*n* = 75)	**Intervention Group** (*n* = 142)
Gross Motor Development	10.87	11.31
Fine Motor Coordination and Visual Motor Integration	12.97	13.13
Emergent Numeracy and Mathematics	9.83	10.24
Cognition and Executive Functioning	7.34	6.63
Emergent Literacy and Language	9.68	9.56
ELOM Total	50.70	50.86

Table 6 displays average ELOM domain scores for the matched children, after removing those who belonged to schools that were not followed up at endline.

Table 6. Average ELOM Scores for the 60 – 70 Month Age Group (Matched).

ELOM Domain	**Comparison Group** (*n* = 32)	**Intervention Group** (*n* = 96)
Gross Motor Development	8.42	11.50
Fine Motor Coordination and Visual Motor Integration	12.60	13.48
Emergent Numeracy and Mathematics	9.62	10.35
Cognition and Executive Functioning	7.70	6.88
Emergent Literacy and Language	9.61	9.74
ELOM Total	*47.94*	*51.94*

The two treatments groups in the older age bracket perform similarly across all domains except for Gross Motor Development (GMD) and Emergent Numeracy and Mathematics (ENM), where the intervention group is Achieving the ELOM Standard and the comparison group is Falling Behind. Given that GMD-related tasks are not included in the BCP's intended work, and not measured by the TBCK, the GMD domain results will not have a significant impact on endline TBCK performance.

To identify whether the matched comparison and intervention groups were significantly different at baseline, an independent samples t-test was conducted. Results concluded that the groups are not statistically different from each other ($t = 1.26$; $p > .05$). This means that the schools that were visited at baseline (and again at endline) are essentially equivalent in nature before and after matching, and that the sample of schools measured at endline are sufficiently equivalent. This ensures that any differences in scores seen at endline can be attributed to children's involvement in BCP, and not to other characteristics of their schools or communities.

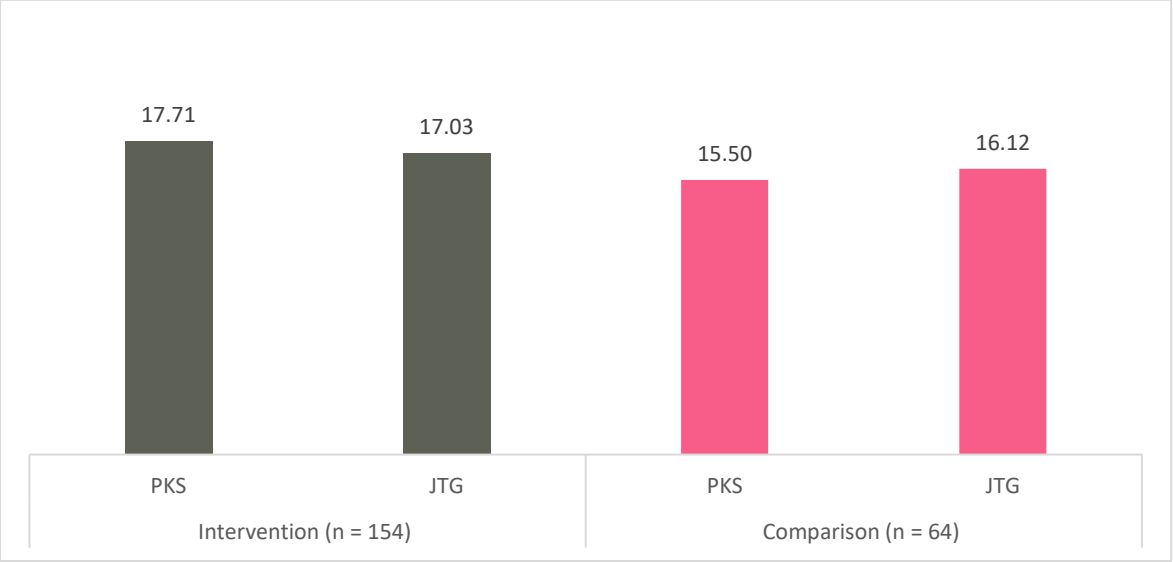
EVALUATION FINDINGS

Endline Findings

Endline TBCK Performance

Figure 7 displays the average TBCK total scores per district, per intervention/comparison group. As can be seen from the figure, the intervention groups in both districts scored higher than the comparison groups. However, these means hide some elements of improvement as measured by the TBCK.

Figure 7. District TBCK Total Mean Scores per Intervention Group.



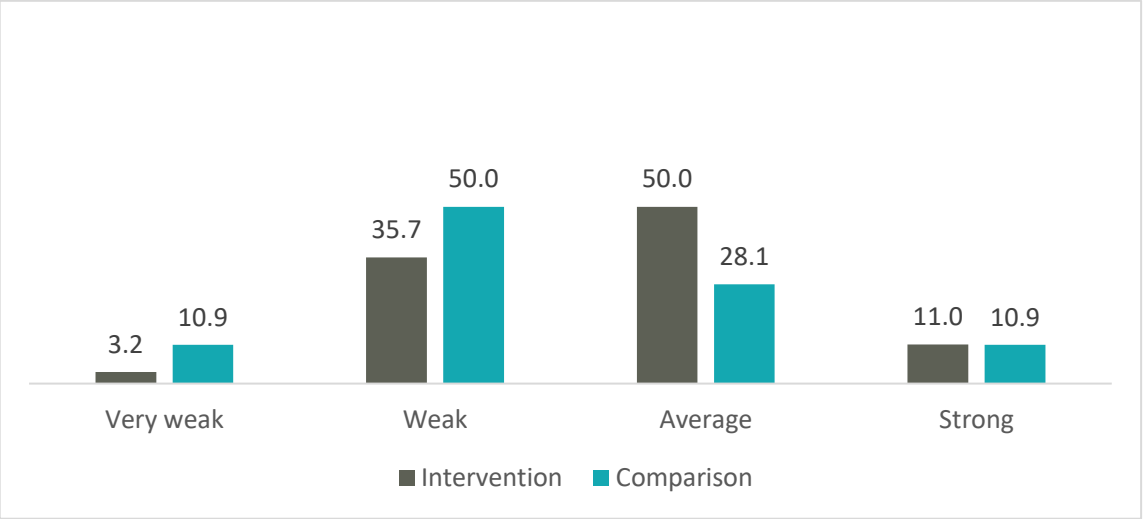
*Note. The maximum score on the TBCK is 24 points.

As noted previously, the TBCK allows learners to be disaggregated into performance bands according to their total scores:

0 - 10	*Very Weak*
11 - 17	*Weak*
18 - 20	*Average*
21 - 24	*Strong*

These performance bands are used to decide whether a child needs intervention or not. Learners in the *weak* and *very weak* bands are not likely to be prepared for the next phase of schooling. BCP aims to address this by moving children into the Average or Strong categories. Figure 8 displays the percentage of learners in each performance band per group. As can be seen in the figure, half of the intervention group fell in the *average* performance band, while half of the comparison group scored in the *weak* band. More than triple the percentage of learners in the *very weak* band belong to the comparison group. However, almost the same percentage of learners from both groups score in the *strong* category. It is possible that this level of preparedness requires other ecological factors, outside of school, to be realised. Alternatively, these results might indicate natural ability of learners from both groups and not be dependent on programme effects. Further, these results may be related to the experience levels of these newly trained mediators, and it might take longer to extend the performance of learners. BCP have, in the past, seen extensions of learners into the higher range of the TBCK. Overall, 61% of BCP learners, compared to 39% of comparison learners, were found to be performing *Average* to *Strong*.

Figure 8. Percentage of Intervention and Comparison Group Children in Each TBCK Performance Band (Matched).



In addition to this descriptive evidence, an independent samples t-test was conducted to assess whether the differences in group-level performance seen in the figures above are statistically significant. On average, learners who received the BCP programme performed better (M = 17.5, SE = 0.24) than learners who did not (M = 15.98, SE =0.51). This difference was significant t (92.13) = -2.68, p < .05. Cohen's d was -0.56, which is indicating a medium effect size. Hence, we can report that learners in the intervention group outperformed learners from the comparison group and this was not down to chance alone, but rather due to the intervention that the learners received. This suggests that children measured during the evaluation, who are taught by a BCP trained teachers, are better prepared to benefit from the next phase of schooling.

Time Series Data (All Data)

Mean TBCK Score

	2007	2008	2009	2010	2011	2012	2015	2016	2018	2020	2022
Comparison		14.76	12.28	13.72		17.90		17.13	15.32	13.05	16.11
Intervention	18.81	18.74	18.51		16.79	18.34	18.76		17.94	17.10	17.54

Percentage Achieving Preparedness (Average/Strong)

	2007	2008	2009	2010	2011	2012	2015	2016	2018	2020	2022	Total
Comparison		36.48%	11.11%	22.20%		60.00%		48.68%	27.86%	6.67%	37.84%	31.35%
Intervention	68.75%	73.53%	72.13%		57.45%	65.93%	80.95%		62.82%	54.48%	61.41%	66.38%

Count Per Year

	2007	2008	2009	2010	2011	2012	2015	2016	2018	2020	2022	Total
Comparison	0	973	144	527	0	20	0	152	140	60	74	2090
Intervention	18	170	290	0	94	273	21	0	156	145	184	1351

When we consider learner performance over the entire timeframe of the running of the Basic Concepts Programme, we find learners who participate in the BCP intervention mirror our findings and perform better than learners who do not – and by a similar margin. On average, learners in the intervention group scored higher (*M* = 18.04, *SE* = 0.1) than learners in the comparison group (*M* = 14.56, *SE* = 0.11). The difference in scores was highly significant *t*(3370)

= -24.05, $p < .001$. Cohen's d was 0.83, which indicates a large effect. Overall, 66% of BCP learners achieve the intended level of preparedness, compared to 31% of comparison schools measured between 2007 and 2022. Covid 19 does not appear to have lessened the effectiveness of BCP.

CONCLUSIONS & RECOMMENDATIONS

This evaluation has established that the Basic Concepts Programme is effective at enhancing the readiness of children taught by BCP teachers in participating schools. The increase in performance is meaningful. Based on the skills measured by the TBCK, children who participate in the programme should be better equipped for the types of thinking required in Grade 1 and beyond.

These results would be impressive from any intervention, but BCP also has the benefit of being scalable. Teachers need only be trained, and to the best of our current knowledge, children taught by trained teachers should continue to benefit from the work of the intervention, even after the teacher leaves the supervision of the programme and continues to use the curriculum.

The success observed by this evaluation is likely partially due to the fact that the intervention operates within schools and capitalizes on a certain degree of guaranteed dosage. Children need only attend school, and they receive the intervention. Children at this age are very likely to attend school, as schools act as childcare for working parents.

We recommend that BCP now look to further scaling the intervention to more teachers. We recommend that funders support this work.

Opportunities for Improvement

We noted that many teachers could not make it through the entire curriculum. The content of the programme includes 6 conceptual domains. This works 6-8 weeks per domain for the first 4 domains, while the 5th and 6th are implemented in the 4th term. Performance on one section of the TBCK (letter) was often lower than other areas measured. This is an area that teachers report not completing in the time that they have with children. It might be that COVID and the way the classes were operating, one day on and one day off, had something to do with the pacing of the programme, and this observation may not extend beyond COVID conditions. However, work should be done to

help teachers make it all the way through the programme in the future.

Further Research

Further research should be done to understand whether children can be moved to the Strong category. This could be done by investigating children who perform in this category, and attempting to understand what makes them different.

Limitations

The conclusions of this evaluation require the schools served by the programme, and the communities that schools serve, to remain relatively stable in terms of socio-economic quintile. Fortunately (or unfortunately) poverty in South Africa, and the social ecology of communities is regarded as relatively stable. Umalusi, the quality assurer for school leaver qualifications makes the same assumptions in adjustments of results. A 5-year norm is commonly employed in this work, and it is expected that communities and educational performance should not change dramatically within this time frame. This is the type of best-practice guidance followed by this evaluation.

Thus, we would not expect schools, previously shown to perform equally, to spontaneously differ in performance. However, we cannot generalize the successes observed in this evaluation to contexts different from those contexts included in our data. We observe that the BCP appears to perform similarly at endline across most communities that BCF have served. We would suggest that additional research be undertaken if the BCP is expanded to communities of socio-economic and language groups not covered by the programmes' history thus far.

APPENDIX

Realised Baseline Sample Characteristics

Table 7. Baseline Age Statistics per Intervention Group (Matched Sample).

	Min Age	Max Age	Average Age	Standard Deviation
Comparison Group	55.9	70.8	64.3	3.5
Intervention Group	55.7	70.9	64.99	3.6
All	**55.7**	**70.9**	**64.8**	**3.6**

The baseline sample was almost evenly split in terms of gender, with 52% of the sample being female, and 48% of the sample being male. The average age of the children in the sample was 64.8 months old, with the youngest child being 55.7 months old, and the oldest child being 70.9 months old. These statistics are presented per intervention group in the table below, showing little difference between the groups.

Figure 9. Baseline Number of Children Stunted per Intervention Group (Matched Sample) – 50 – 59 Months.

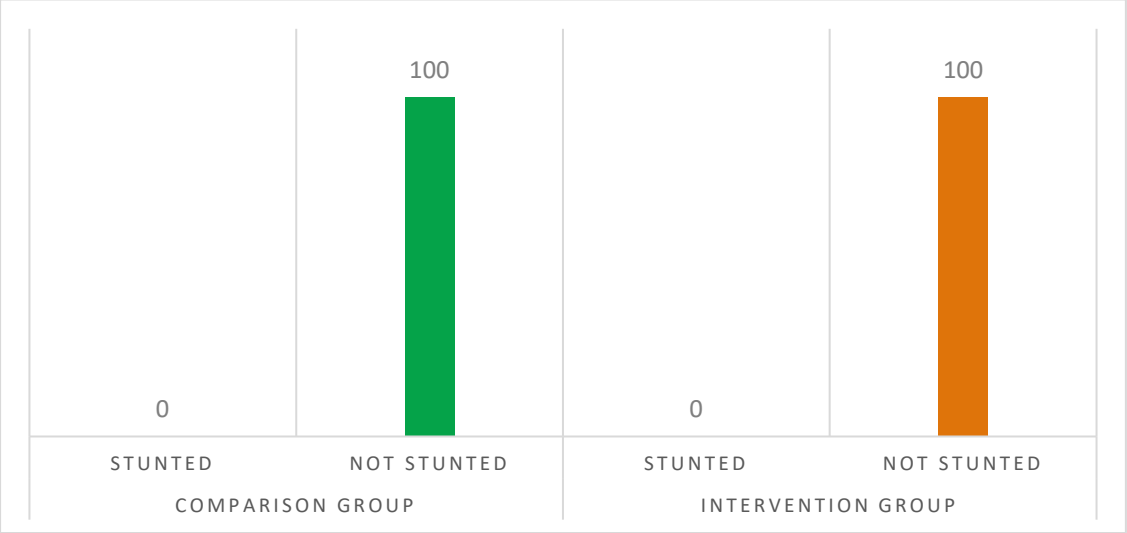
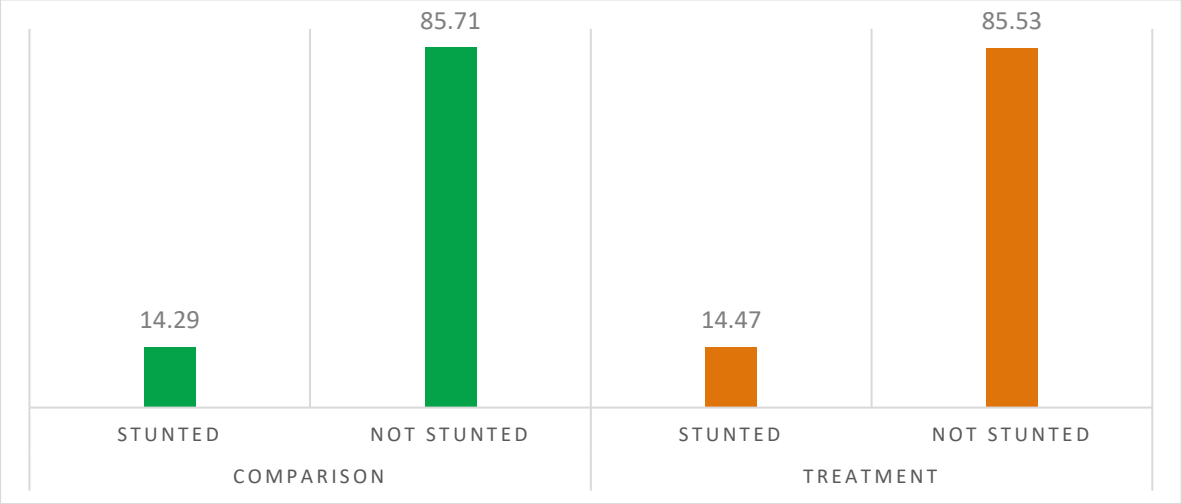


Figure 10. Baseline Number of Children Stunted per Intervention Group (Matched Sample) – 60 – 69 Months.



The majority of the matched baseline evaluation sample (76%) receive the Child Support Grant. More children in the matched intervention group receive the grant (n = 77; 71.96%) compared to the matched comparison group (n = 32; 88.89%). However, given that the comparison sample is one third of the size, these two are proportionately the same.

Figure 11. Baseline Number of Children in Receipt of the Child Support Grant (Matched Sample).