Developing Teachers' Competence in Teaching Mental Mathematics in Grades 2 and 3

REPORT SUBMITTED TO RTI INTERNATIONAL BY THE WITS MATHS CONNECT PRIMARY TEAM

NOVEMBER 1, 2024

Prepared by Sameera Hansa, Shemunyenge Taleiko Hamukwaya, Kate Sehowa, and Anthony A. Essien
University of Witwatersrand, Johannesburg, South Africa

Acknowledgments

Our profound gratitude goes to the teachers and students who participated in this research and development initiative, especially the teachers who willingly gave us access to their classrooms for observations. Our thanks also go to Refilwe Maropola for assisting with data collection, and Lawan Abdulhamid for his assistance with the statistical analysis of data.

List of Acronyms

ANOVA	analysis of variance
BTT	base-ten thinking
BTTG2	Base-Ten Thinking Grade 2
MSAP	Mental Starters Assessment Project

Table of Contents

Executive Summary 4
1. Introduction 5
2. Project Model 7
2.1 Design 7
2.2 Population and Sampling 8
2.3 Instrumentation 9
3. The Intervention 12
3.1 Grade 2 BTT 12
3.2 Grade 3 MSAP 13
4. Training of Teachers, Departmental Heads, and Subject Advisors 13
5. Data Collection 15
5.1 Pre- and Post-Tests 15
5.2 Classroom Observation 16
6. Data Capturing and Analysis 16
7. Validity through Cross-Checking Data 18
8. Ethical Considerations 18
9. Findings from the Study 19
9.1 Results from the Quantitative Analysis 20
9.2 Implementation Fidelity and Its Impact on Performance 35
10. Conclusion 42
11. References 44
Appendix 1: Test Items – Jump Strategy 46
Appendix 2: Test Items – Bridge Strategy 48
Appendix 3: Test Items – BTT 50
Appendix 4: Implementation Fidelity 52

Executive Summary

This project aimed to enhance teachers' efficacy in teaching (and grade 2 and 3 students' proficiency in understanding) mental mathematics through the base-ten thinking (BTT) framework, addressing persistent challenges in early grade mathematics education in South Africa. In South Africa, as in many other countries, student performance in mathematics remains low and is yet to reach the expected level. This low performance is often compounded by inefficient strategy use—evident in both teaching and learning practices—which hinders students' progression to efficient calculation strategies and negatively affects their ability to work within higher number ranges.

Grounded in Morrison's (2018) doctoral research, the BTT framework integrates constructs of Freudenthal's theory of number structuring and mental mathematics strategies such as the jump and bridging through ten strategies. These approaches aim to develop students' number sense and foster efficient mental calculation.

A total of 12,423 students participated, with 1,416 in the control group and the remainder in the experimental group. Using a design research model, the study combined quantitative pre- and post-test analysis with qualitative insights from video-recorded lessons. Teachers in the experimental group received training on BTT pedagogical content knowledge, lesson plans, and instructional resources to support early grade mental mathematics. Guidelines on administering the pre- and post-tests were also provided.

The findings reveal a positive, statistically significant impact of BTT mental starter strategies on students' numerical understanding. Grade 2 students in the experimental group demonstrated the largest gains (12.6 percentage points), while grade 3 students showed smaller but meaningful gains in the jump strategy (2.3 points) and bridging through ten strategy (4.1 points). Improvements in the control group, particularly in the jump and bridge strategies, were attributed to general classroom learning rather than the intervention. Item-level analysis, particularly in items that tested mathematical reasoning and application of the BTT strategies, highlighted the intervention's effectiveness in improving mental math skills.

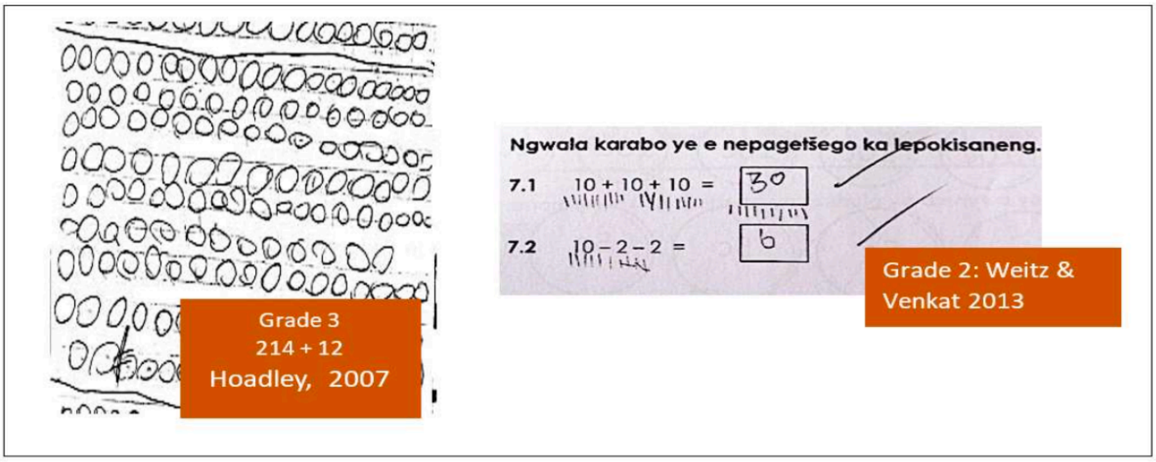
In terms of implementation fidelity, our study found that high-fidelity schools tended to demonstrate stronger gains in student performance, particularly in jump strategy tasks, compared to schools with moderate fidelity. This pattern suggests a potential correlation between implementation fidelity and student achievement in mental mathematics tasks.

These results demonstrate the potential of targeted, evidence-based interventions to address foundational gaps in early mathematics education. By enhancing students' mental mathematics skills through BTT, this initiative can contribute to long-term improvements in mathematics outcomes across South African classrooms. Continued refinement of instructional strategies and robust teacher training are essential for sustained success.

1. Introduction

A broad consensus arising out of a number of research projects in South Africa is that the overarching problem with early number learning is linked to the persistent use of counting in ones in many public schools. As an example, analysis by Hoadley (2007) of the scripts of students in a high-stakes assessment at the end of grade 3 revealed that for the question 214 + 12 = __, many of the students attempted the question through the count-all strategy (that is, they drew 214 objects and 12 more objects and then counted all objects in an attempt to arrive at the answer). A similar situation is reported by Weitz and Venkat (2013), as shown in Figure 1.

Figure 1. Prevalence of one-by-one counting and lack of move to counting with structure



This inefficient strategy—evident in both teaching and learning practices—hampers students' progression to more efficient calculating strategies and negatively affects their facility with working in higher number ranges. One consequence of this over-reliance on counting in ones in the foundation phase (grades 1–3) is that when students move into the intermediate phase (grades 4–6) and beyond, they have a poor grasp of the base-ten numeration system and the related understanding of place value. In fact, research into the state of primary education in South Africa has indicated that a good number of students fall behind by as much as two to three years below their actual grade by the time they are in grade 6 (see Hartley, 2007; Spaull & Kotze, 2015; Mohohlwane & Taylor, 2015; Human et al., 2015). It is evident that this cumulative growth in the knowledge gap starts in the early grades (see, for example, research by Fritz et al., 2020). While these learning gaps cannot be attributed to one factor alone, they raise questions about the teacher's role in creating meaningful opportunities for learning mathematics in early grade classrooms in South Africa and the extent to which teachers teach for understanding in these grades. In this vein, Fullan's (1993, p. 5) anecdote that teacher education has the honor of being simultaneously the worst problem and the best solution in education comes to mind (see Essien, Matthews, et al., 2023).

Base-ten thinking (BTT) as a framework speaks to using 10 as a reference point, using 10 (and multiples of 10) as a unit when solving multi-digit tasks, and organizing numbers and calculations into 1s, 10s, and 100s for ease of calculation. Using the BTT framework, we upskilled teachers and departmental heads on how to make judicious use of the first 15 minutes that are allocated for mental mathematics in the early grades in the South African curriculum, with the aim of better supporting mathematics teaching and learning on the ground and improving numeracy in the foundation phase (early grades). Our effort to develop teachers' competence in teaching mental mathematics in grade 2 is called Base-Ten Thinking Grade 2 (BTTG2), and for grade 3 it is called the Mental Starters Assessment Project (MSAP). The BTTG2 acts as a precursor to the MSAP. The key goal of both the BTTG2 and MSAP is to equip mathematics departmental heads and in-service mathematics teachers with the skills and knowledge needed for working with the base-ten structure of numbers for efficient calculation.

The BTT framework is based on findings from a study by Morrison (2018) that focused on developing early number skills in a context of widespread low student attainment in numbers using Wright et al.'s (2006) Mathematics Recovery program for grouped intervention. Morrison's study used a sample of 20 grade 2 students (ten each for intervention and control groups) who were considered middle attainers in their cohort. Intervention students were equitably split into two quartets (four per group) and two "singletons" based on test results and pragmatic reasons. Both groups and singletons were withdrawn from class twice a week for 18 intervention sessions of 40 minutes each, which totaled 12 hours for each group and each singleton. The intervention used a "test-teach-test" model consisting of individual task-based interviews using Mathematics Recovery assessments as a pre-test, then 18 grouped/individual intervention sessions, and then a repeat of the task-based interviews as the post-test. Students' individual task-based interviews were video-recorded, transcribed, and analyzed using descriptors from the Learning Framework in Number (Wright et al., 2006). The study found a positive correlation between enhanced learning outcomes and the implementation of BTT among grade 2 students.

In 2021–22, the Wits Maths Connect Primary project embarked on a research and development project aimed at strengthening teachers' ability to teach number sense through structuring numbers and calculations using the BTT framework. As Mason, Stevens, and Watson (2009) and Mulligan and Mitchelmore (2009) argue, structure permeates the whole of mathematics at every stage and for every age and is of critical importance in developing mathematics competence in children. Research has shown that young children who understand structure in mathematics acquire a deep conceptual understanding and perform better than those with an underlying poor understanding of mathematical structure (see, for example, Mulligan, 2002). Moreover, research has indicated that the ability to reason about number relations during the primary school years is a more accurate predictor of future mathematical achievement than proficiency in arithmetic procedures (Nunes et al., 2009, in Askew, Graven & Venkat, 2022).

In 2023, the project entered its refinement phase, continuing to work with the same cohort of schools to maintain continuity in training and intervention. This phase focused on fine-tuning and enhancing the intervention based on areas of improvement identified during the pilot phase. The iterative nature of design research enabled focused adjustments in teacher training, materials delivery, and overall fidelity of implementation across participating schools. Qualitative insights from teacher feedback and classroom observations, together with quantitative data from student assessments (pre- and post-tests), were used to guide these refinements in the project's design and inform further adjustments.

In 2024, a revised version of the teacher and student books was introduced. In addition, we adopted a scaled-up model of training subject advisors and departmental heads with the aim that they would then train the teachers in their respective schools. However, after the first training session in February, it became evident that the intended cascade model was not being consistently implemented. The training of teachers did not always occur, and, as a result, we had to step back and train teachers directly in the second training.

The project objectives are as follows:

- To develop and explore the capacity of mathematics teachers and departmental heads to support grade 2 and 3 teachers in their own schools in implementing the tasks and resources package aimed at building base-ten thinking among students.
- To investigate learning gains through pre- and post-testing of students in the classes of participating teachers compared to non-participating teachers (control group).

The research questions for this project are as follows:

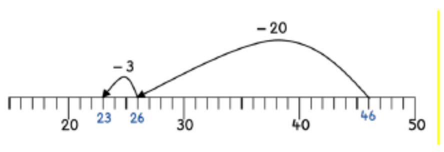
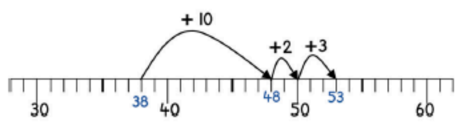
1. What learning gains are made in early grade classes taught by teachers who are part of a professional development initiative where base-ten thinking strategies are used to develop teacher attention to number structure when teaching mental mathematics?
2. How do teachers who are part of a professional development initiative where base-ten thinking strategies are used implement these strategies in their classrooms? To what extent is there fidelity (or not) to the mathematics examples contained in the student book (grade 2) and teacher book (grade 3)?

2. Project Model

2.1 Design

As indicated earlier, two key additive strategies based on BTT are central to the project model: the jump strategy and the bridging through ten strategy. Both strategies leverage a number line as a visual tool, keeping the first number whole while adding or subtracting in place value chunks. This promotes not only greater accuracy and efficiency in calculations but also a better understanding of number relationships and relative sizes. These strategies are illustrated through number sentences and represented on semi-structured number lines.

Figure 2. Examples of jump and bridging through ten strategies

Jump Strategy	Bridging Through Ten
46 – 23 = ☐	38 + 15 = ☐
46 – 20 – 3 = 23	38 + 10 + 2 + 3 = 53
	

The project focuses on enhancing grade 2 students' mental mathematics skills, making it a valuable initiative within the South African context. A key implementation approach is to start with the number line and then progressively fade the number line so that the solution becomes completely mental.

The MSAP introduces similar strategies, beginning with the jump strategy and bridging through ten, using the more advanced empty number line image, and expands to cover four additional mental strategies: doubling and halving, rounding and adjusting, re-ordering, and linking addition and subtraction. Like the BTTG2 effort, the MSAP aims to shift students away from inefficient counting strategies such as finger counting or tally marks. These time-consuming methods can lead to errors (especially as the number range increases) and hinder the development of a robust number sense.

2.2 Population and Sampling

The participants in this study were sampled from the Ugu District in KwaZulu Natal Province, South Africa. We worked with 60 foundation phase teachers and departmental heads identified by the district (based on our criteria) in each grade 2 and 3 class (totaling 120 teachers for both the BTTG2 and MSAP), with approximately 40 students per class. The 120 teachers were selected by district officials based on our request to have different categories of schools (rural, urban, semi-urban, mixed/multi-grades) present in the study.

Additionally, four control schools were included to assess the intervention's effectiveness for the BTTG2, and five control schools were included to assess the intervention's effectiveness for the MSAP. All schools received information about the study and the data collection process, with consent obtained from principals, teachers, parents and guardians, and students. We asked the district to also assist us in identifying control schools, and both the control and the experimental schools completed pre- and post-tests to measure outcomes.

Table 1. Total number of students expected to participate in the project

	Grade 2	Grade 3
Experimental school	3,000	4,700
Control school	350	203

2.3 Instrumentation

Two primary instruments were used for data collection in this study: written tests (pre- and post-tests) and video-recorded classroom observations with a sample of participant teachers. Figures 3–4 show examples from the written tests. For the MSAP, the focus was on two key strategies: the bridging through ten strategy and the jump strategy, as mentioned previously.

Figure 3. Sample of grade 2 assessment: Base-ten strategy

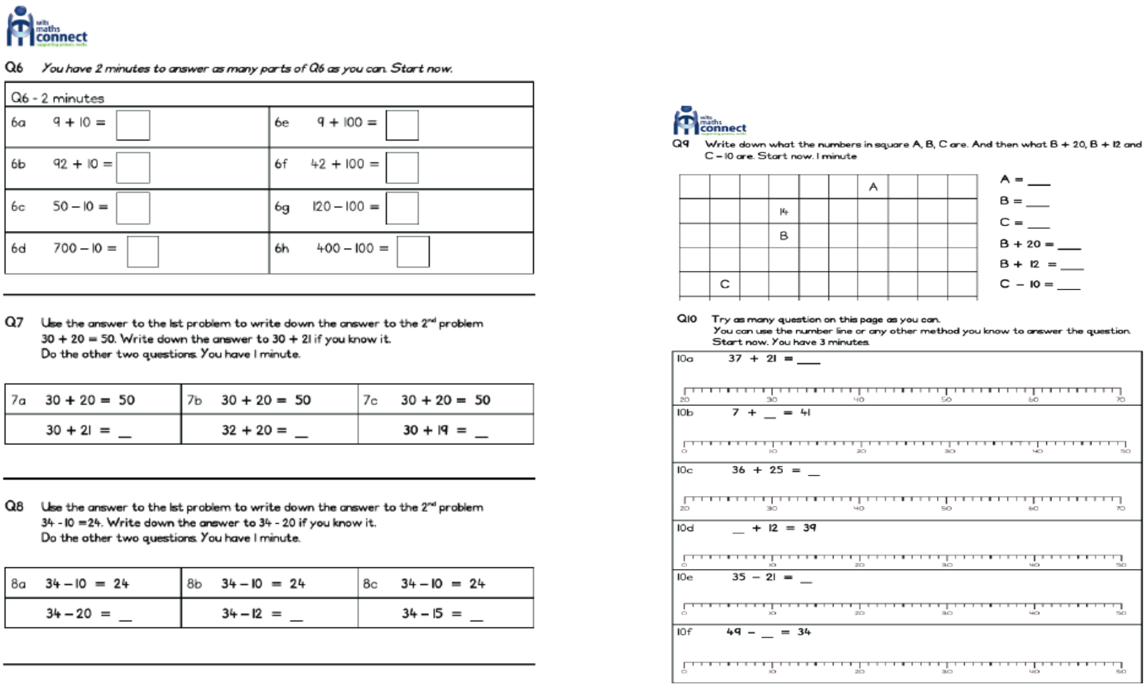


Figure 4. Sample of grade 3 MSAP assessment: Bridge strategy

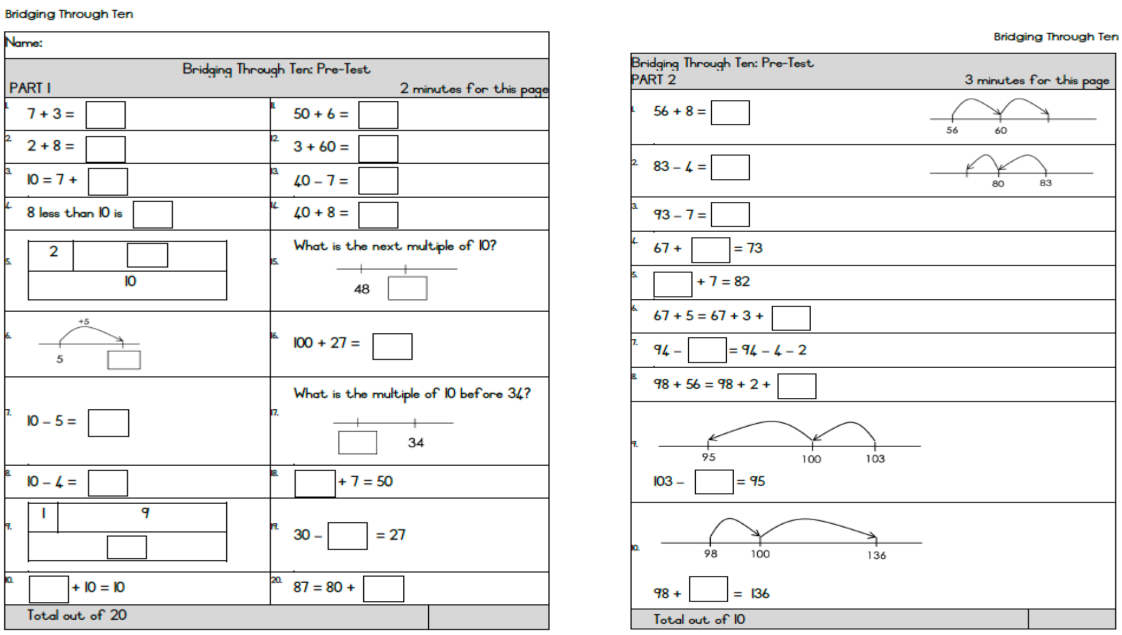
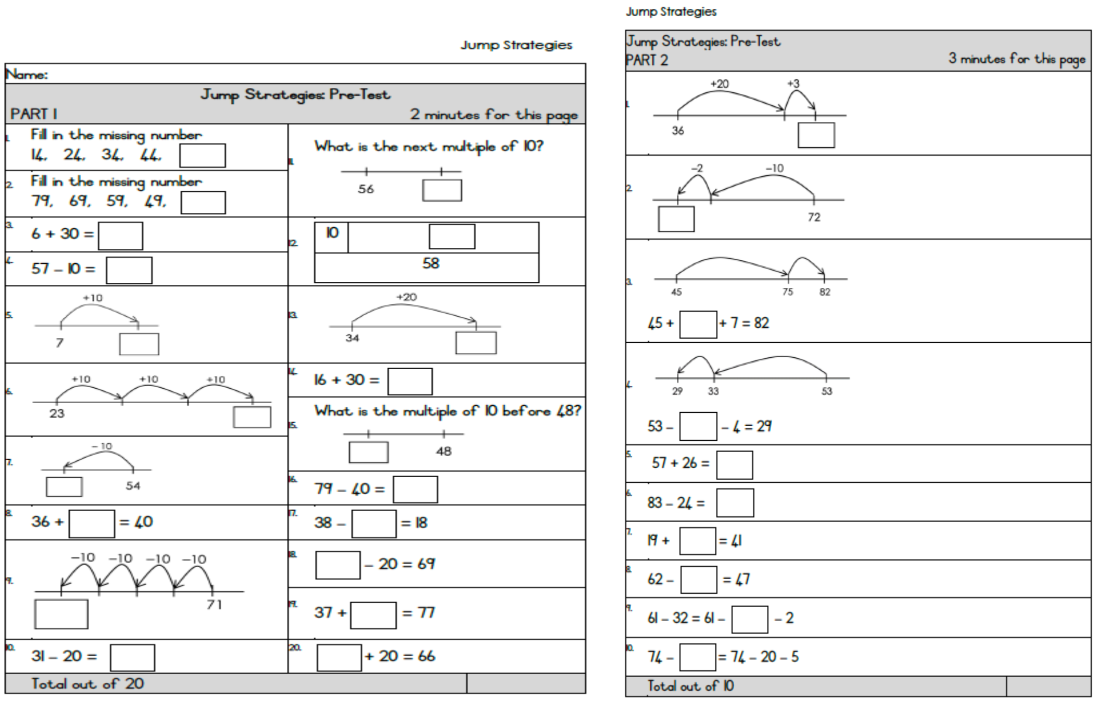


Figure 5. Sample of grade 3 MSAP assessment: Jump strategy



2.3.1 Base-Ten Thinking Grade 2

The BTTG2 portion of the project is based on a “test-teach-test” model. The written test—administered as both a pre-test and a post-test—consists of ten timed sections, each containing a set of questions designed to discourage students from counting in ones (see Appendix 1). Conducted in a whole-class setting, this teacher-led assessment typically takes 30–40 minutes to complete. A scripted teacher instruction sheet provides explicit guidelines for timing and delivering each question set, ensuring standardized administration across classrooms. Students respond on answer sheets that align with the teacher’s prompts, with different colors used to distinguish between pre-test and post-test scripts. The primary aim of the written test is to establish a broad understanding of students’ early number knowledge immediately before and after the intervention period.

2.3.2 Mental Starters Assessment Project

The MSAP also follows a “test-teach-test” model, consisting of six mental mathematics lesson starter units for grade 3 students. Each unit emphasizes a distinct calculation strategy aligned with the curriculum, focusing on a specific set of interconnected skills. Each unit is three weeks long, and the structure of each unit is as follows:

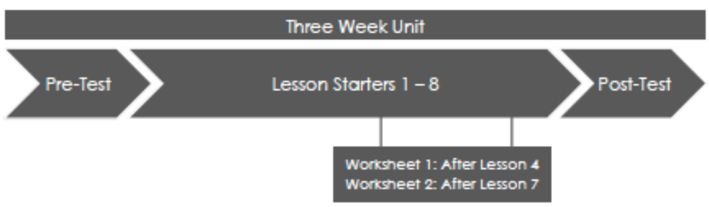


Figure 6 below shows the units that are covered in the first three terms of the school year. For this year's data analysis and reporting, the focus is specifically on the first two strategies: bridging through ten and jump.

Figure 6. Different strategies used in the MSAP

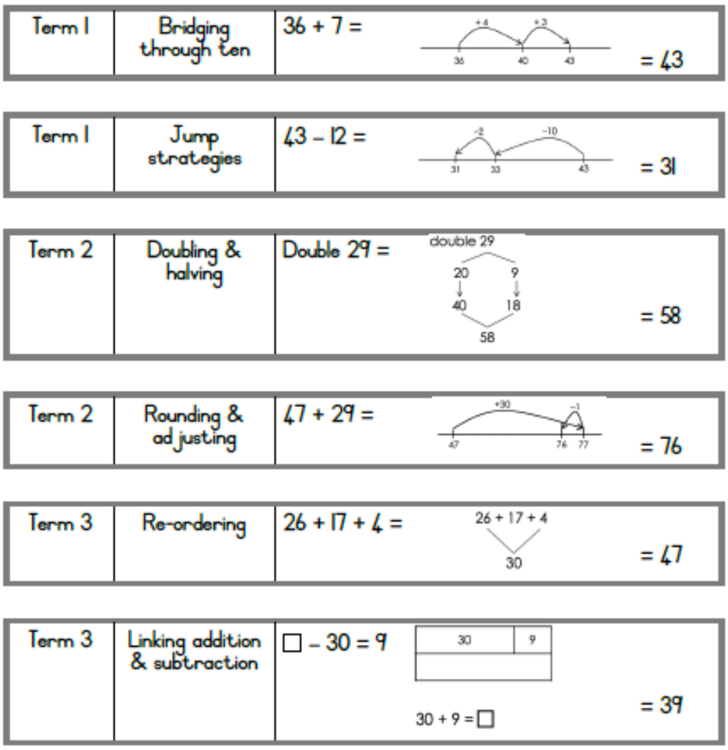
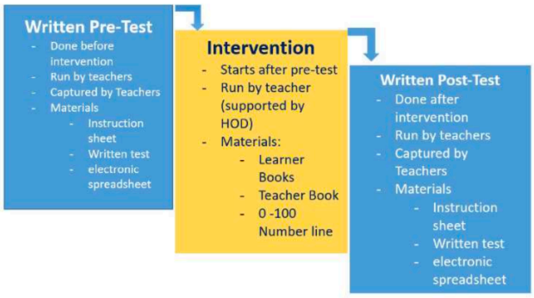


Figure 7 below provides an overview of the intervention model for grades 2 and 3.

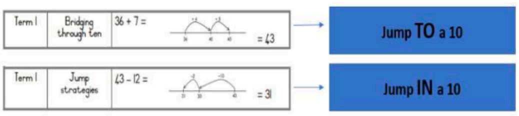
Figure 7. Overview of the intervention model

Grade 2 Base Ten Thinking Programme

Intervention Model

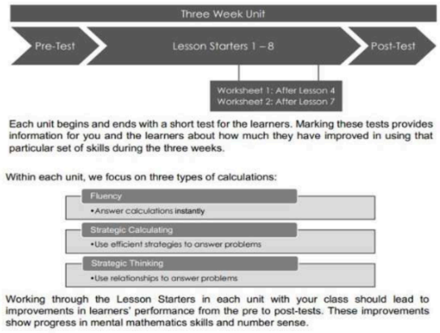


Only 2 Calculation Strategies Covered

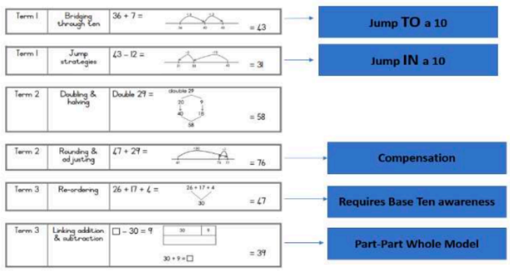


Grade 3 Mental Starts Assessment Project

Intervention Model



Extention to 6 strategies



In both models:

- Lesson starters are designed to fit into the 15-minute oral and mental mathematics at the start of grades 2 and 3 mathematics lessons.
- Each lesson starter starts with a one-minute mental warm-up focused on important rapid recall skills for that unit. Students are expected to answer these questions quickly and confidently.
- A lesson starter task sequence follows the warm-up. The task sequence in each lesson starter explains how to teach the strategy and what to write on the board.

After each lesson starter, there are individual tasks for students to complete.

3. The Intervention

3.1 Grade 2 BTT

The intervention (or "teach") component of the BTTG2 model involves grade 2 teachers and students using the BTT resources for the 15-minute mental mathematics segment of the daily numeracy lesson, four days a week, over a total of eight weeks in each of terms 1, 2, and 3. The teacher book includes four mental math lesson plans per week for eight weeks, alongside the corresponding worksheets found in the student workbook. Teachers are expected to teach one mental mathematics lesson each day for 15 minutes, with the fifth day dedicated to consolidation activities. Each student's workbook contains a worksheet for every lesson, aligned with the lessons presented in the teacher book. The teacher book is written in English, while the student workbooks are translated into the language of learning and teaching (in the case of the Ugu district, IsiZulu). This approach is designed to meet the curriculum expectations in South Africa and to facilitate meaning-making among students.

3.2 Grade 3 MSAP

The MSAP model is similar to the BTTG2 model, as it also employs a test-teach-test framework. The "teach" component involves grade 3 teachers and students using the MSAP resources during the 15-minute mental mathematics segment of the daily numeracy lesson throughout terms 1, 2, and 3. The teacher book contains six units, which are each three weeks long. Each unit begins and concludes with a brief assessment for students, allowing for the evaluation of their progress in mastering the specific skills covered during that period. Each lesson starter commences with a one-minute mental warm-up focused on essential rapid recall skills relevant to the unit, encouraging students to respond quickly and confidently. Following the warm-up, a task sequence is presented, detailing the instructional approach for teaching the strategy and guidance on what to write on the board. After completing each lesson starter, students engage in individual tasks designed to reinforce the newly acquired strategy, with an emphasis on mental computation rather than counting in ones. If students encounter difficulties, they are encouraged to demonstrate their reasoning—for example, by using an empty number line or bar diagram. These sketches should be informal, as the objective is to facilitate rapid mental processing. Each lesson starter plan includes a link to a brief video illustrating the strategy in action.

4. Training of Teachers, Departmental Heads, and Subject Advisors

During the planning phase, we established a schedule of four training days (in a year) for departmental heads and subject advisors—two sessions before the project commenced and two additional sessions midway through the year. In the first training session, held in February 2024, departmental heads and subject advisors were trained with the intention that they would subsequently train their respective teachers, thereby reaching a broader cohort of students. However, it soon became evident that the anticipated cascade/scale-up model was not effective, as not all departmental heads and subject advisors returned to train the teachers under their supervision. Consequently, in the second training session, we extended invitations to include the teachers as well.

Table 2. BTTG2 training sessions

Training session	**Areas covered**
1	• Rationale for the project (why BTT) • Overview of the project • Need for different types of testing and the benefits thereof • How to teach the base-ten tasks for terms 1 and 2 • How to use the materials for mental math sessions • How to conduct the test (materials handed out)
2	• Recapping the rationale for the project and different types of assessment • How to teach the base-ten tasks for terms 1 and 2 • Pedagogic training through micro-teaching certain tasks • Hand out written tests, teacher instruction sheets, and an electronic copy of the spreadsheet for how to capture the written test results • Hand out teacher books and number lines per school
3	• Recap on the rationale for the project (why BTT) and an overview of the project • How to teach the base-ten tasks for terms 3 and 4 • How to use the materials for mental math sessions • How to conduct the test (materials handed out) • Pedagogic training through micro-teaching certain tasks
4	• How to teach the base-ten tasks for terms 3 and 4 • Pedagogic training through micro-teaching certain tasks • Feedback on pre-test results • How to conduct the post-test and capture students' results • Get teachers to volunteer to be observed in their classrooms (video-recorded) • Hand out consent forms

Table 3. MSAP training sessions

Training session	Areas covered
1	• Rationale and overview of the MSAP • Need for different types of testing and the benefits thereof • How to teach the base-ten tasks for each of the units for terms 1 and 2 • Pedagogic training through micro-teaching certain tasks • How to use the materials for mental math sessions • How to conduct the test (materials handed out) • How to conduct the pre-test and capture students' results
2	• Recap the rationale and overview of the MSAP • How to teach the base-ten tasks for each of the units for terms 1, 2, and 3 • Pedagogic training through micro-teaching certain tasks • How to use the materials for mental math sessions • How to conduct the test (materials handed out) • Feedback on pre-test results • How to conduct the post-test and capture students' results • Get teachers to volunteer to be observed in their classrooms (video-recorded) • Hand out consent forms

An essential component of the support structure for teachers involved the appointment of two part-time coaches. Their responsibilities extended beyond providing routine support: they actively engaged in conducting regular classroom visits to observe instructional practices and offer tailored feedback to teachers and to the project team. This in-person coaching not only helped reinforce the pedagogical strategies introduced in professional development sessions but also provided teachers with real-time assistance in addressing challenges as they arose within their classrooms when it comes to the implementation of the project materials and other related project matters.

5. Data Collection

5.1 Pre- and Post-Tests

At the outset of the program, a written pre-test was administered by grade 2 and grade 3 teachers across all participating schools, encompassing both experimental and control groups. This pre-test provided the baseline data for the study. To assess the impact of the program, a corresponding post-test was conducted across the same cohort of schools by mid-October 2024. This timing allowed for a comparison of student performance before and after the implementation of the program, offering insights into the effectiveness of the intervention across different school environments.

5.2 Classroom Observation

In August 2024, the project team conducted a series of video-recorded classroom observations in the district to document instructional practices among teachers who had provided consent for recording. Over a five-day period, the team visited 11 schools and observed 12 teachers, with six of the schools receiving three visits. These recorded lessons focused on the teachers implementing instructional strategies from the program. To comply with ethical standards, only students who had consented to be filmed were included in the recordings, ensuring respect for participant privacy and informed consent protocols.

Table 4. List of video-recorded classroom observations

Schools	Grade	Number of visits		Number of teachers
N	3	3	**School visits x 3**	1
B	3	3		1
E(2)	3	3		2
Mb	3	1		1
S	3	1		1
M	2	1		1
It	2	1		1
En	2	1		1
Es	2	3	**School visits x 3**	1
No	2	3		1
I	2	3		1
			Total	**12**

6. Data Capturing and Analysis

Data collected and captured for both the pre- and post-test in both grades for each cohort is presented in Tables 5 and 6.

Table 5. Captured data for experimental group

	Pre-test	Post-test	Total
Grade 2 BTT	1,998	1,649	3,647
Grade 3 MSAP jump strategy	2,213	2,121	4,334
Grade 3 MSAP bridge strategy	1,958	2,484	4,442

Table 6. Captured data for control group

	Pre-test	Post-test	Total
Grade 2 BTT	256	248	504
Grade 3 MSAP jump strategy	183	230	413
Grade 3 MSAP bridge strategy	256	243	499

In the process of analyzing the assessment data, a V-LOOKUP function was employed in Microsoft Excel to facilitate the matching of students who participated in both the pre-test and post-test. This method aimed to create a comprehensive dataset that accurately reflected the performance trajectories of individual students over the course of the intervention. However, while matching data were successfully retrieved for most students who completed both tasks, certain inconsistencies were noted within the recorded results. Specifically, some students were identified solely by their last names, while others used different first names in the pre-test compared to their post-test submissions. These discrepancies posed significant challenges in ensuring the accuracy of the matching process. Furthermore, it was observed that some test scripts lacked the identification of the schools attended by the students, which further complicated efforts to associate the results with specific schools. These challenges highlight the importance of standardized data collection practices in educational assessments to facilitate effective analysis and reporting.

We obtained the following matched data based on students who took both the pre- and post-test for each cohort, as presented in Tables 7 and 8.

Table 7. Matched data for experimental group (pre- and post-test)

	Total
Grade 2 BTT	392
Grade 3 MSAP jump strategy	787
Grade 3 MSAP bridge strategy	693

Table 8. Matched data for control group (pre- and post-test)

	Total
Grade 2 BTT	157
Grade 3 MSAP jump strategy	148
Grade 3 MSAP bridge strategy	159

Apart from descriptive statistics, we used Sigma-XL to run statistical tests (particularly paired t-tests) to determine the statistical significance of our results. These results are presented and interpreted as part of the quantitative data analysis. In addition to this, we also conducted an analysis of variance (ANOVA) to examine the difference in mean performances from pre- to post-test for the experimental and control groups. We used the means from the pre-tests to map experimental schools with control schools. Thereafter, comparisons in the mean between these schools were made. For grade 2 BTT, we used the means from schools that obtained 47% or lower, whereas for the grade 3 MSAP results, we looked at the bottom five performing schools in the pre-test and mapped them to the five control schools.

Our final level of analysis addresses the issue of implementation fidelity, aiming to respond to the second part of research question 2: *To what extent is there fidelity (or not) to the mathematics examples contained in the student book (grade 2) and teacher book (grade 3)?* Specifically, this analysis examines the extent to which the intervention strategies are implemented as intended, in alignment with the guidelines provided in the teacher book. Consequently, from the 24 videos recorded, we selected 15 videos from schools that underwent three observations each: three videos from grade 2 and twelve from grade 3. Each video has an approximate average duration of 20 minutes.

7. Validity through Cross-Checking Data

To ensure the validity of the findings in this report, a robust cross-checking process was employed to verify the integrity and consistency of the data collected and captured from the pre-test and post-test. In this report, cross-checking involved systematically comparing results from the pre-test and post-test assessments to identify any discrepancies or inconsistencies in students' data. This process included matching student identifiers across both tests to ensure that capturing was done effectively. According to Johnson and Christensen (2014), such methodological triangulation enhances the reliability of findings, thereby reinforcing the overall validity of the results.

Furthermore, discrepancies identified during the cross-checking process, such as variations in student names or missing school identifiers, were meticulously reviewed and resolved wherever possible. This meticulous approach aligns with best practices outlined by Creswell (2014), who emphasizes the importance of addressing data integrity issues to bolster the credibility of research outcomes. By employing a comprehensive cross-checking mechanism, the report not only safeguarded against potential biases but also ensured that the findings were reflective of genuine changes in students' performance.

8. Ethical Considerations

Ethical clearance for this research was secured from the Ethics Committee of the University of the Witwatersrand, as well as from the Provincial Department of Basic Education given that we worked with public schools. Prior to engaging with the participating schools, permission was obtained from the district directors responsible for overseeing all public schools in the district. Furthermore, the principals and teachers were duly informed about the nature of the project, enabling them to make informed decisions regarding their participation. Consent forms were subsequently distributed to those principals and teachers who opted to take part in the study. Additionally, consent was obtained from the parents or guardians of the students involved in the program.

9. Findings from the Study

The questions that our research and development project set out to answer were as follows:

1. What learning gains are made in early grade classes taught by teachers who are part of a professional development initiative where base-ten thinking strategies are used to develop teacher knowledge about attention to number structure when teaching mental mathematics?
2. How do teachers who are part of a professional development initiative where base-ten thinking strategies are used implement these strategies in their classrooms? To what extent is there fidelity (or not) to the mathematics examples contained in the student book (grade 2) and teacher book (grade 3)?

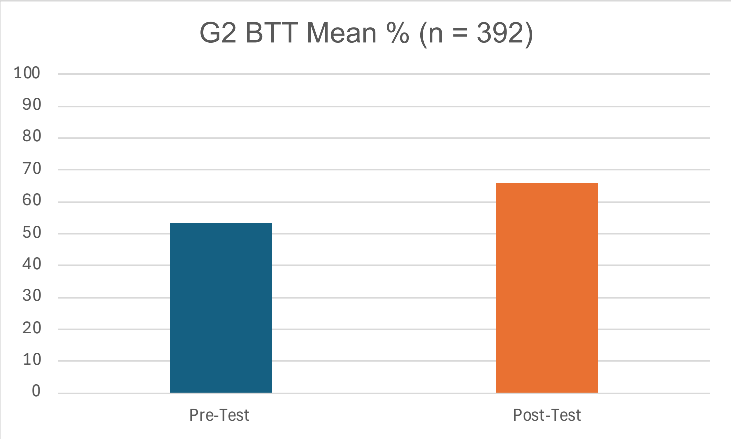
To answer these questions, we first provide an analysis of the overall performance shifts from pre- to post- test for both the experimental and the control group. We then engage with item facility (performance) shifts for both groups. Given that there were only four control schools in grade 2 and five control schools in grade 3, we matched the results of these schools with those of the experimental group in terms of achievement in the pre-test. In doing this, we chose equivalent number of classes (control vs. experimental) for statistical analysis. Finally, to answer research question 2, we provide a comparison of experimental schools and control schools in terms of fidelity of implementation and impact thereof on performance. In doing the latter, we also engage with how the teachers implement the mental mathematics strategies that they were inducted into.

9.1 Results from the Quantitative Analysis

9.1.1 Overview of Experimental Group's Performance in the Pre- and Post-Tests

Grade 2 base-ten thinking

Figure 8. Pre- and post-test results for grade 2



The graph above illustrates the mean percentage scores of grade 2 students' BTT performances before and after the intervention, with data matched from a sample of 392 students. In the pre-test, students achieved a mean score of just over 50% (53.3%). Following the intervention, the post-test mean score increased to 65.9%, suggesting a possible positive impact of the instructional approach on students' understanding of BTT. To determine whether this difference was statistically significant, a paired t-test was conducted using Sigma-XL (see Table 9).

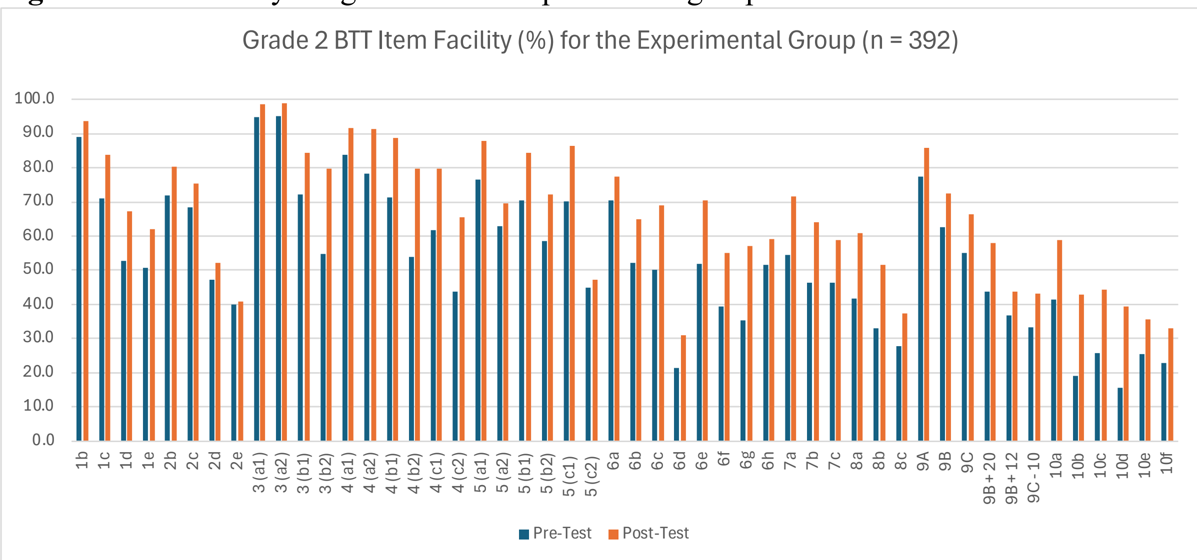
The paired t-test results show a mean difference of 12.597 between post-test and pre-test scores, with a standard deviation of 22.830. The calculated t-value is 10.925, and the p-value for this two-sided test is < 0.0001. Given that the p-value is significantly below the 0.05 threshold, we reject the null hypothesis, indicating a *statistically significant* improvement in post-test scores compared to pre-test scores. The 95% confidence interval for the mean difference ranges from 10.330 to 14.864, further supporting that the observed difference is unlikely to be due to chance. This suggests that the intervention likely had a positive impact on students' BTT skills.

Table 9. Paired t-test results (pre-test/post-test) for grade 2 BTT experimental group

Results	**Post-test/pre-test**
Count	392
Mean	12.597
StDev	22.830
SE mean	1.153
t	10.925
P-value (2-sided)	**0.0000**
UC (2-sided, 95%)	14.864
LC (2-sided, 95%)	10.330

Item analysis

Figure 9. Item facility for grade 2 BTT experimental group



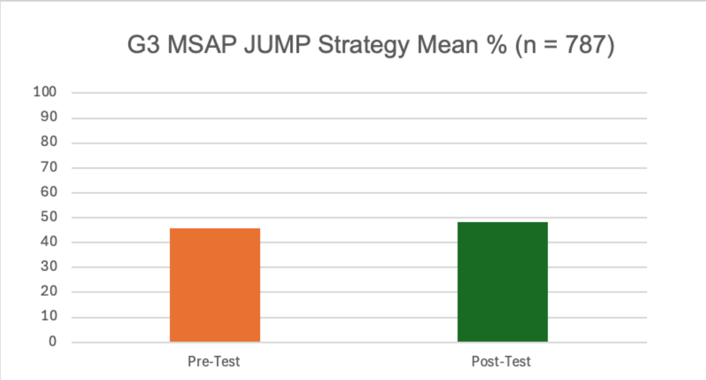
The graph above presents a comparison of pre-test and post-test scores for matched students who participated in the BTTG2 intervention (sample size N=392). This visualization allows us to assess the learning gains associated with the use of the BTT strategies, based on students' performances across 50 different test items (see Appendix). On all of the items, the post-test scores are higher than the pre-test scores, indicating a general trend of improvement in performance following the BTT intervention. This suggests that the intervention may have been effective in helping students enhance their mental math skills.

The magnitude of score increase varies across items. Some items, particularly items 3a1 and 3a2, exhibit high pre-test scores and show marginal improvement. These items required students to write down the numbers 9 and 89 and to also write the number that is *one more than* each of these numbers. The facilities in these items indicated that students found these items relatively easy. In contrast, other items, especially test items 10a, b, and d (which are missing addend questions) show lower pre-test scores (mostly below 40%) but show an improvement of about 20% in the post-test. These items were more challenging and provide a good indication that the BTT strategies taught in the intervention may have contributed to a better grasp post-intervention.

Overall, the BTTG2 intervention appears to have positively influenced student outcomes, as evidenced by higher post-test scores across most items, supported by the paired t-test result.

Grade 3 MSAP jump strategy

Figure 10. Pre- and post-test results for grade 3 jump strategy



The graph above shows the mean percentage scores for grade 3 students' performance in the MSAP jump strategy based on pre-test and post-test results of a matched sample of 787 students. The mean score in the pre-test was 45.8%, while the post-test mean increased to 48.1%. The shift from pre- to post-test indicates an improvement in students' understanding of the jump strategy. To determine whether this shift was significantly different, a paired t-test was conducted using Sigma-XL.

The paired t-test analysis provides a statistical analysis of the difference between pre-test and post-test scores for the MSAP jump strategy (see Table 10). The mean difference between post-test and pre-test scores is 0.685, with a standard deviation of 5.850. The test statistic (t) is 3.284, and the p-value for this two-sided test is 0.0011. Since the p-value is well below the 0.05 threshold, we reject the null hypothesis, indicating a *statistically significant* difference between the pre-test and post-test scores.

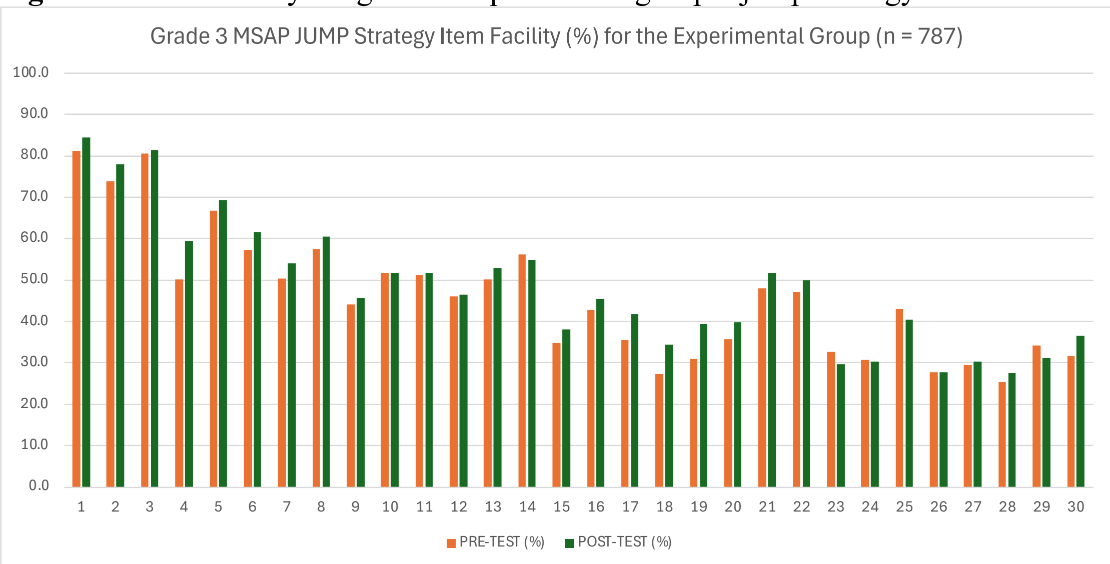
The 95% confidence interval for the mean difference ranges from 0.276 to 1.094, suggesting a small but statistically significant improvement in scores. Although the increase is modest, it provides evidence that the MSAP jump strategy intervention had a measurable, positive effect on students' scores.

Table 10. Paired t-test results (pre-test/post-test) for grade 3 MSAP jump strategy

Results	Post-test/pre-test
Count	787
Mean	0.684879
StDev	5.850
SE mean	0.208534
t	3.284
P-value (2-sided)	**0.0011**
UC (2-sided, 95%)	1.09422821
LC (2-sided, 95%)	0.275530

Item analysis

Figure 11. Item facility for grade 3 experimental group – jump strategy



The graph above presents a comparison of pre-test and post-test scores in the jump strategy test (sample size N=787). This visualization allows us to assess the learning gains associated with the use of the jump strategy, based on students' performance across 30 different test items. Across all the items, the post-test scores are higher than the pre-test scores, indicating a general trend of improvement in performance following the jump strategy intervention. This suggests that the intervention was effective in helping students enhance their mental math skills.

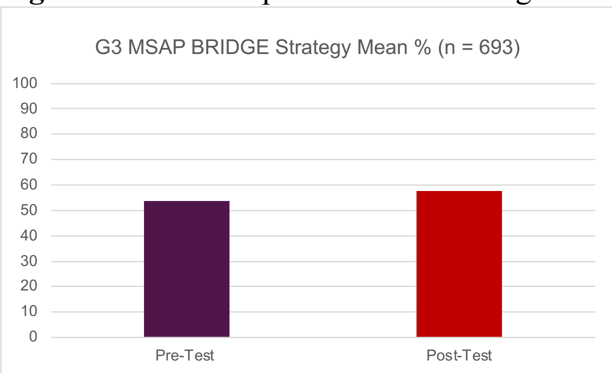
The magnitude of score varies across items. Some items, particularly items 1, 2, and 5 (see the graph above), exhibit high pre-test scores (around 70–80%) and show marginal improvement. These items required students to fill in the missing number in a sequence (items 1 and 2) and add 10 using the number line representation (item 5), indicating that students found these items relatively easy before the intervention. In contrast, items such as

17, 18, 19, and 25–30 (missing number calculations) show lower pre-test scores (mostly below 40%) but show some improvement post-test. These items represent more challenging concepts where the jump strategy intervention may have contributed to a better grasp, albeit still at a lower mastery level than other items.

The differences in performance gains across items indicate variability in how students benefited from the strategy. While some items show only a slight increase or no improvement, others show more marked gains, suggesting that the strategy may have been more effective for certain types of mental mathematics problems.

Grade 3 MSAP bridge strategy

Figure 12. Pre- and post-test results for grade 3 bridging through ten strategy



The graph above shows the mean percentage scores for grade 3 students' performance in the MSAP bridge strategy, based on pre-and post-test results for 693 matched students. The pre-test mean score was 53.7%, and after the intervention, the post-test mean increased to 57.8%. This improvement of 4.1 percentage points suggests a possible positive effect of the intervention on students' understanding or application of the bridge strategy. To determine whether this shift was significantly different, a paired t-test was conducted using Sigma-XL (see Table 11).

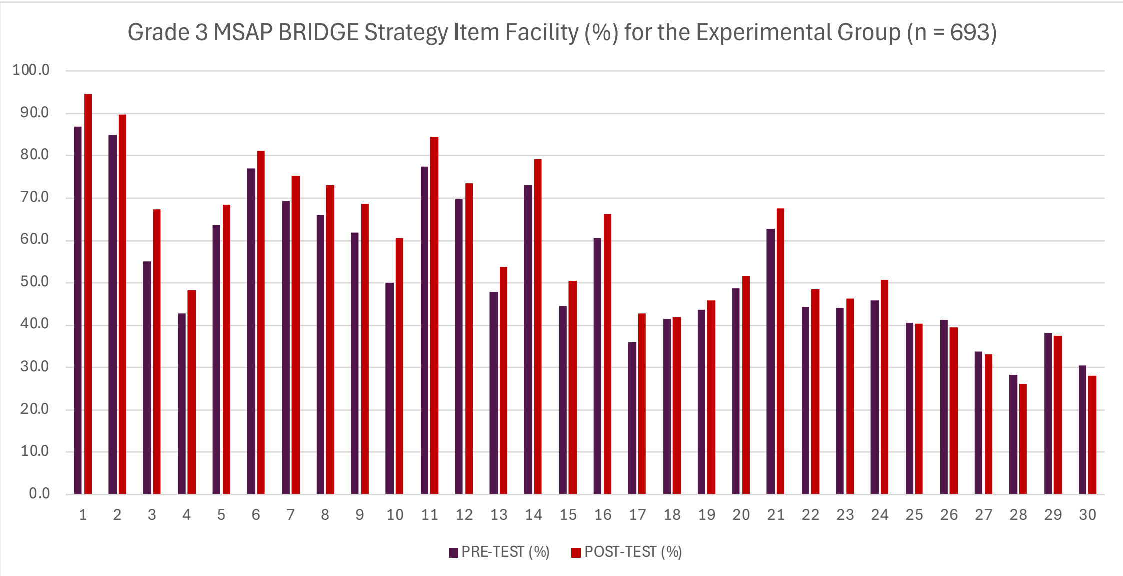
The paired t-test results show a mean difference of 1.229, with a standard deviation of 5.620. The test statistic (t) is 5.759, and the two-sided p-value is < 0.0001, which is far below the 0.05 significance level. This result leads us to reject the null hypothesis, indicating a *statistically significant* improvement in post-test scores compared to pre-test scores. The 95% confidence interval for the mean difference, ranging from 0.810 to 1.649, suggests that the observed increase is unlikely to be due to chance. This result implies that the intervention possibly had a positive and statistically significant impact, albeit with a modest effect size, on participants' post-test scores.

Table 11. Paired t-test results (pre-test/post-test) for grade 3 MSAP bridge strategy

Results	Post-test/pre-test
Count	693
Mean	1.229
StDev	5.620
SE mean	0.213487
t	5.759
P-value (2-sided)	**0.0000**
UC (2-sided, 95%)	1.649
LC (2-sided, 95%)	0.810278

Item analysis

Figure 13. Item facility for grade 3 experimental group – bridging through ten strategy



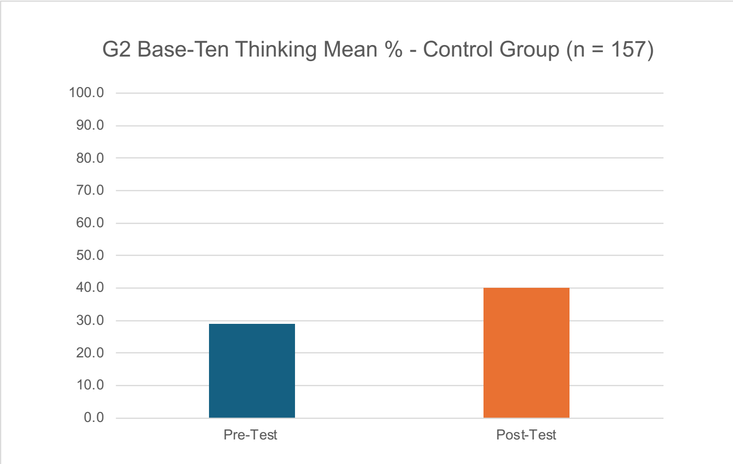
The data presented in the graph above compare the pre-test and post-test scores of students using the MSAP bridge strategy for mental mathematics, with a sample size of 693. The graph visually presents the percentage scores across 30 different mathematics-test items for both the pre-test and post-test, showing the possible impact of the intervention on student performance.

Across most of the test items, post-test scores are higher than pre-test scores, albeit varying in magnitude. Some items show substantial gains (e.g., test items 1, 5, 11, and 21), while others exhibit more modest increases or relatively similar pre- and post-test percentages (e.g., items 27–30). Although items 26, 28, and 30 showed a negative gain, the overall performance shows a marginal improvement, with the paired t-test suggesting that the intervention on the bridge strategy may have contributed to the overall performance shifts in the post-test.

9.1.2 Overview of Control Group's Performance in the Pre- and Post-Tests

Grade 2 BTT

Figure 14. Pre- and post-test results for grade 2 control group



The results from the pre-test and post-test for the grade 2 control group show an improvement in BTT. The mean percentage score increased from 29.0% in the pre-test to 40.0% in the post-test, based on a sample size of 157 students. This indicates a positive shift in the control group's understanding of base-ten concepts over the testing period, suggesting that some factors, potentially including general classroom instruction or time, may have contributed to this improvement. To determine whether this shift was significantly different, a paired t-test was conducted using Sigma-XL (see Table 12).

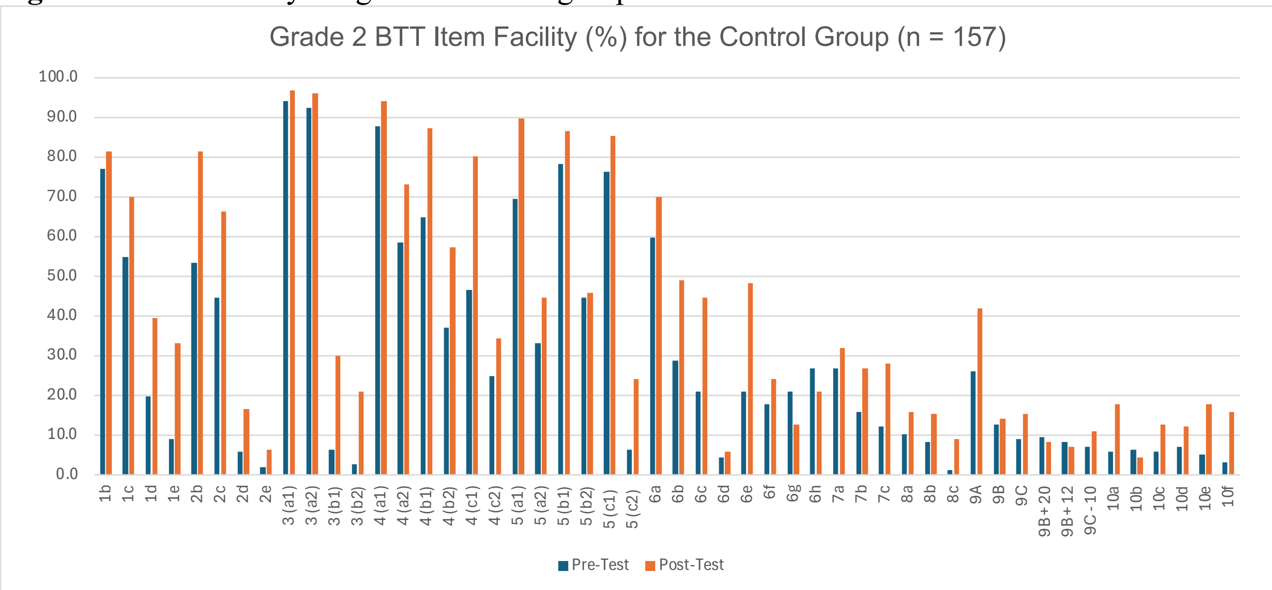
The t-value of 21.427 and a two-sided p-value of 0.0000 indicate that the improvement from pre-test to post-test is statistically significant. The 95% confidence interval for the mean difference ranges from 26.702 to 32.126, suggesting that the true mean improvement in scores lies within this range. These results provide evidence that the observed increase in BTT scores is not due to random variation, indicating a meaningful improvement of base-ten concepts over the testing period.

Table 12. Paired t-test results (pre-test/post-test) for grade 2 control group

Results	Post-test/pre-test
Count	157
Mean	29.414
StDev	17.201
SE mean	1.373
t	21.427
P-value (2-sided)	**0.0000**
UC (2-sided, 95%)	32.126
LC (2-sided, 95%)	26.702

Item analysis

Figure 15. Item facility for grade 2 control group

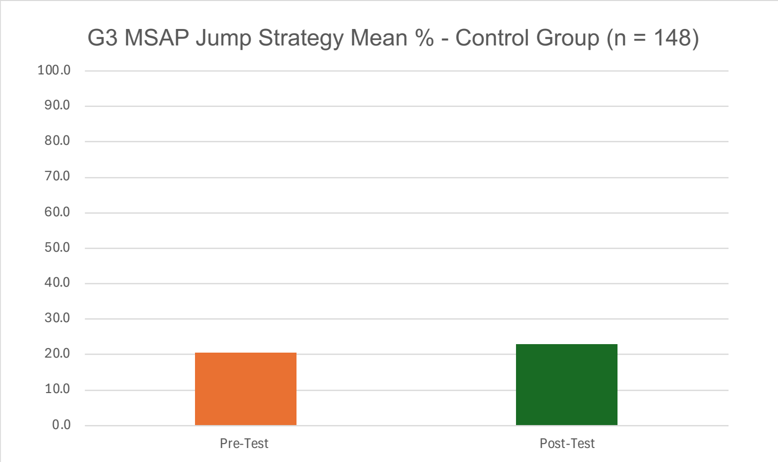


The graph above shows varying levels of initial performance across the items in the pre-test. High initial scores are observed on items 3a1, 4a1, and 5a1, with item facility percentages of 94.3%, 87.9%, and 69.4%, respectively, indicating that these items were relatively easier for students. However, items like 2e, 4c1, and 3b2 scored very low on the pre-test, with facility percentages as low as 1.9%, suggesting that these items posed more difficulty for students initially. In the post-test, there is a noticeable improvement in facility percentages across most items, given that students wrote the pre and post-test in February and October 2024. Large increase gains were noticed in items such as 1d, 1e, and 2b, and in question 4. However, certain items, such as 6g (-8.3%) and 6h (-5.7%), showed negative gains, indicating a drop in performance on these items in the post-test compared to the pre-test.

Lastly, in questions 7–10, which tested mathematical reasoning and application of BTT skills, student performance remained low despite the gains made from pre- to post-test. With the exception of question 9A, facilities across these items were still around 30% or less. The gains made in these items can be attributed general classroom learning and still indicate a stronger need for intervention on BTT skills, which would enhance overall mental mathematics skills.

Grade 3 MSAP jump strategy

Figure 16. Pre- and post-test results for grade 3 control group – jump strategy



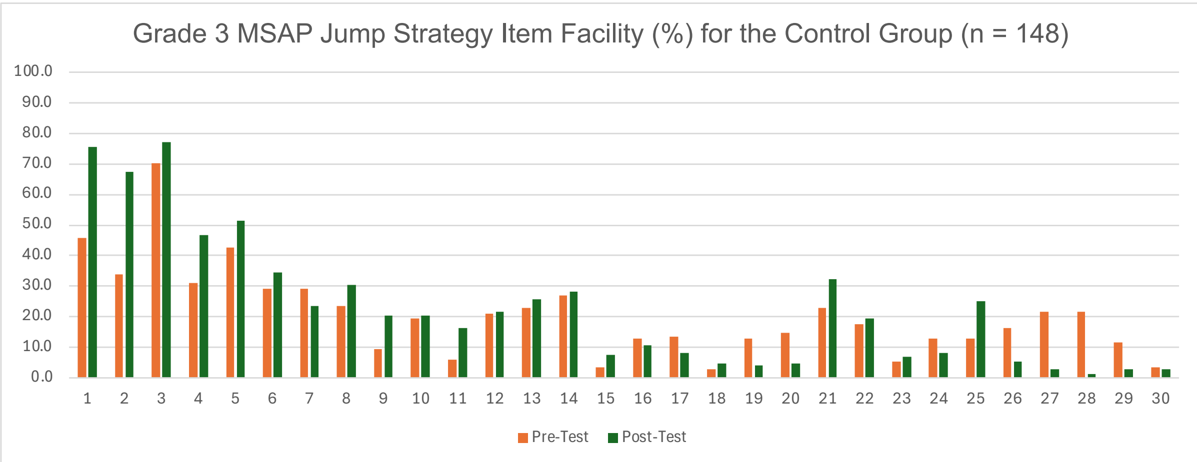
The pre-test and post-test results for the grade 3 MSAP jump strategy in the control group (N=148) show a slight increase in the mean percentage score. The mean score rose from 20.6% on the pre-test to 22.9% on the post-test. This improvement indicates a marginal gain in the group's understanding or application of the jump strategy. To determine whether this increase is statistically significant, a paired t-test was run on Sigma-XL. The paired t-test showed a t-value of 1.550, and a two-sided p-value of 0.1233, which is above the common significance threshold of 0.05 (see Table 13). The 95% confidence interval for the mean difference ranges from -0.626 to 5.175. Since the confidence interval includes zero and the p-value is > 0.05, the result *is not statistically significant*. Therefore, we can conclude that there is no evidence of a meaningful change in scores from pre- to post-test.

Table 13. Paired t-test results (pre-test/post-test) for grade 3 MSAP jump strategy control group

Results	Post-test/pre-test
Count	148
Mean	2.275
StDev	17.856
SE mean	1.468
t	1.550
P-value (2-sided)	0.1233
UC (2-sided, 95%)	5.175
LC (2-sided, 95%)	-0.625803

Item analysis

Figure 17. Item facility for grade 3 control group – jump strategy



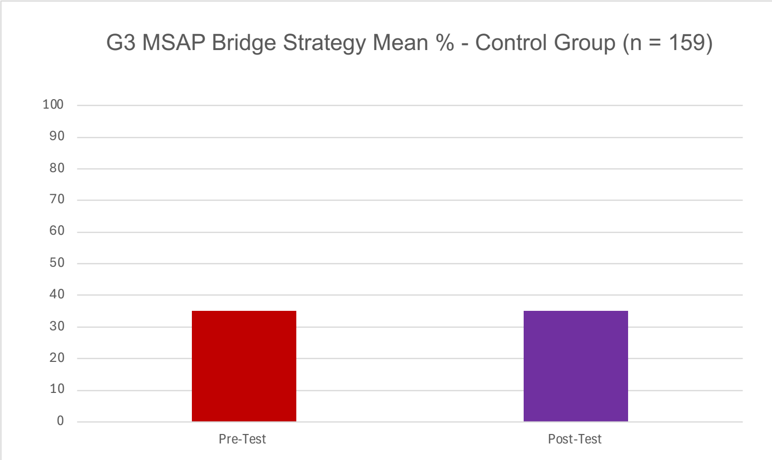
The pre-test scores display a range of initial proficiencies across the 30 items. Items such as item 3 (6 + 30), with a facility percentage of 70.3%, and items 1 (fill in the missing number) and 5 (7 + 10 = __), with 45.9% and 42.6%, respectively, indicate areas where students demonstrated moderate initial understanding. However, certain items, including items 9, 15, 18, and 23, had notably low pre-test scores, with item facility percentages as low as 6.1%, 3.4%, and 3.4%, suggesting that these items were particularly challenging for students. These items, together with finding the missing addend and subtrahend in part two of the test, present challenges for students in both the pre- and post-test.

In the post-test, notable increases were observed in items 1 and 2, which rose from 45.9% to 75.7%, and from 33.8% to 67.6%, respectively. While these items showed gains, the level of difficulty is low as they required students to fill in the missing number in a sequence of given numbers.

Conversely, certain items exhibited minimal or negative gains. For instance, item 18 showed little improvement (2.7% in the pre-test and 4.7% in the post-test), while items 7, 16, 17, 19, 20, 24, and 26–30 noted decreasing facilities in the post-test. This trend indicates a strong need for further instructional focus and practice with the jump strategy, which would provide students with the skills to answer questions of this nature.

Grade 3 MSAP bridge strategy

Figure 18. Pre- and post-test results for grade 3 MSAP – control group



The pre-test and post-test results for the grade 3 MSAP bridge strategy in the control group, consisting of 159 matched participants, reveal a mean percentage score of 35% in both the pre- and post-test, indicating no observed change in the control group's performance. This stable mean score suggests that the control group, which did not receive the intervention, maintained the same level of performance over the testing period. To determine whether this shift is statistically significant, a paired t-test was conducted on Sigma-XL (see Table 14).

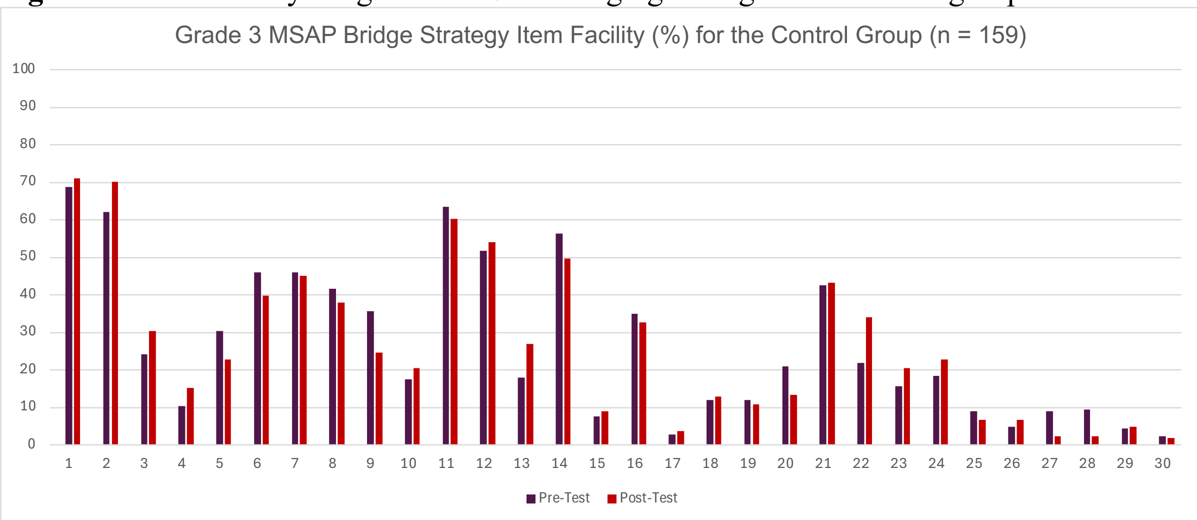
The paired t-test showed a t-value statistic of -0.125473 and a two-sided p-value of 0.9003, which suggests that there is no statistically significant difference between post-test and pre-test scores. The 95% confidence interval for the mean difference ranges from -2.808 to -2.472, which includes zero, therefore aligning with the non-significant p-value, indicating no meaningful difference. The very small mean difference and wide confidence interval suggest that any observed change is likely due to random variation rather than a real effect.

Table 14. Paired t-test results (pre-test/post-test) for grade 3 MSAP bridge strategy control group

Results	Post-test/pre-test
Count	159
Mean	-0.167715
StDev	16.855
SE mean	1.337
t	-0.125473
P-value (2-sided)	0.9003
UC (2-sided, 95%)	2.472
LC (2-sided, 95%)	-2.808

Item analysis

Figure 19. Item facility for grade 3 MSAP bridging through ten – control group



The graph above demonstrates the pre and post-test results of the bridge strategy for the grade 3 control group. The pre-test results show a broad range of initial proficiencies. High pre-test scores were observed on items such as item 1 (6 + 4 = __) (68.7%) and item 11 (50 + 7 = __) (63.5%), indicating that a significant number of students had prior familiarity with the skills assessed in these questions. In contrast, several items had much lower pre-test scores, including item 4 (8 less than 10) (10.4%), item 17 (what is the multiple of 10 before 34?) (2.8%), and items 25–30 (missing number problems) (ranging from 2.4% to 9%). The low facilities in these items suggest that these items were challenging for the grade 3 students.

The post-test results indicate some variations in student performance after the assessment period. Some items showed a slight improvement from pre- to post-test, while other items reflected limited gains or decreasing facilities.

Upon examining changes from pre-test to post-test, certain items showed slight improvements, while others experienced decreases. Notable decreases occurred for items 5 and 9, where facility percentages dropped substantially. Items 5 and 9 were based on the part-whole concept and required students to find the missing part or whole.

Overall, a few items demonstrated modest gains, but the overall low gains across items suggest that the bridging through ten strategy remained challenging for the control group, especially on items requiring advanced application of the strategy.

The lower pre-test scores observed in the control schools, despite random selection, may be attributed to several contextual and situational factors. Variability in the quality of the teachers' content knowledge could have contributed to disparities in the pre-test performance. Additionally, socioeconomic differences between schools, such as varying levels of parental involvement, access to resources, or community challenges, may have impacted student readiness. Cohort-specific characteristics, including a higher proportion of students with learning difficulties or language barriers in the control schools, could also explain the discrepancies. Furthermore, factors such as school infrastructure, class sizes, or access to learning materials (student books) may have influenced pre-test results. While random selection reduces systematic bias, natural variability between schools can still result in pre-test results differences. A deeper analysis of school-level data could provide further insights into these disparities.

9.1.3 Comparison of Performance between Experimental Group and Control Group

To assess whether the gains between the experimental group and control group were statistically significant, a one-way ANOVA test was conducted.

For the grade 2 dataset, the analysis focused on 14 experimental schools with pre-test mean scores below 48%, compared to four control schools with pre-test mean scores of 48% or less.

In the grade 3 dataset, the focus shifted to the jump and bridge strategies, comparing the lowest-performing five schools in the experimental group during the pre-test with five control schools during the post-test.

- *Jump strategy:* The five experimental group schools had pre-test mean scores ranging from 20% to 40%, while the control group schools had pre-test mean scores between 20% and 33%.
- *Bridge strategy*: The five experimental group schools recorded pre-test mean scores between 40% and 48%, compared to the control group schools, which had pre-test mean scores ranging from 23% to 45%.

This approach allowed for a nuanced comparison of performance across both strategies and groups.

Grade 2 BTT

Figure 20. Trends in performance from pre to post-test between the experimental group and control group for grade 2 BTT

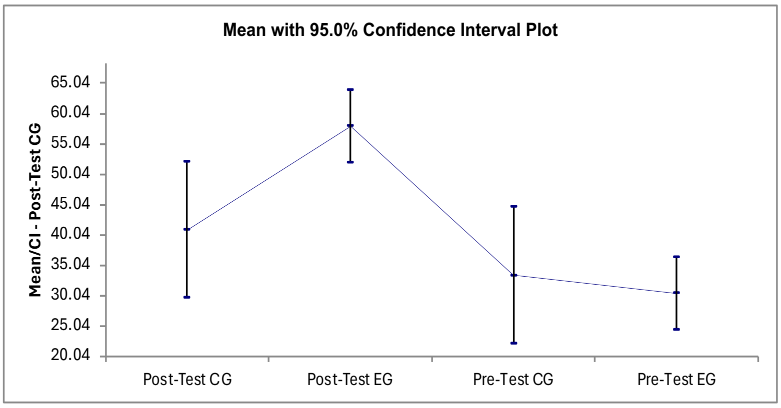


Table 15. ANOVA results for grade 2 BTT

Source	SS	DF	MS	F	P-value
Between	5723.7603	3.0000	1907.9201	15.632	**0.0000**
Within	3905.6700	32.0000	122.0522		
Total	9629.4303	35.0000			

The results indicate that both the control group and experimental group improved from pre-test to post-test, with the experimental group showing a substantially higher post-test mean score (see Figure 20). The confidence interval for the experimental group is also higher than that of the control group, highlighting greater improvement in the experimental group.

The ANOVA results presented in Table 15 show a highly significant p-value ($p < 0.05$), confirming a statistically significant difference in post-test scores between the groups. The experimental group's higher post-test mean (57.94) compared to the control group's (40.91) suggests that the intervention had a meaningful positive effect on performance. This improvement in the experimental group is likely due to the intervention rather than chance.

Grade 3 MSAP jump strategy

Figure 21. Trends in performance from pre to post-test between the experimental group and control group for grade 3 MSAP jump strategy

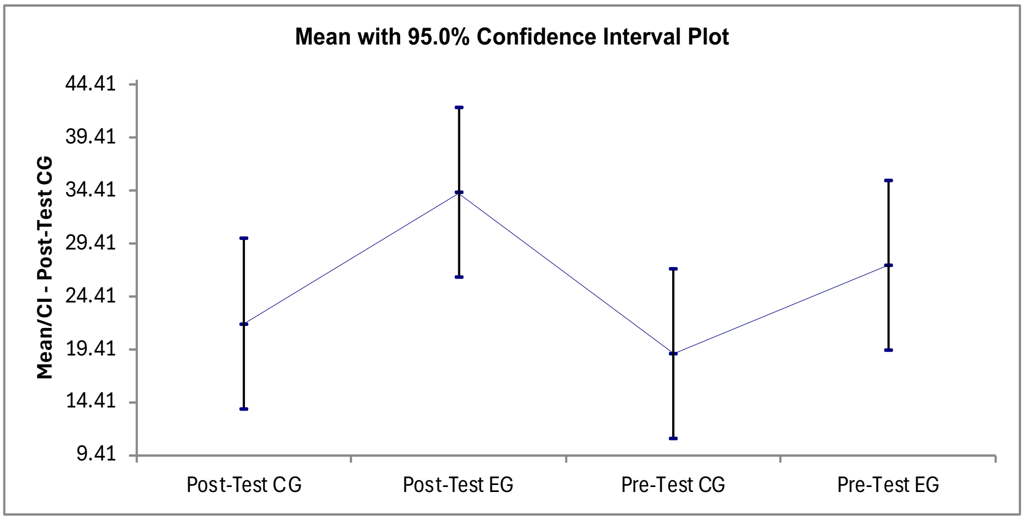


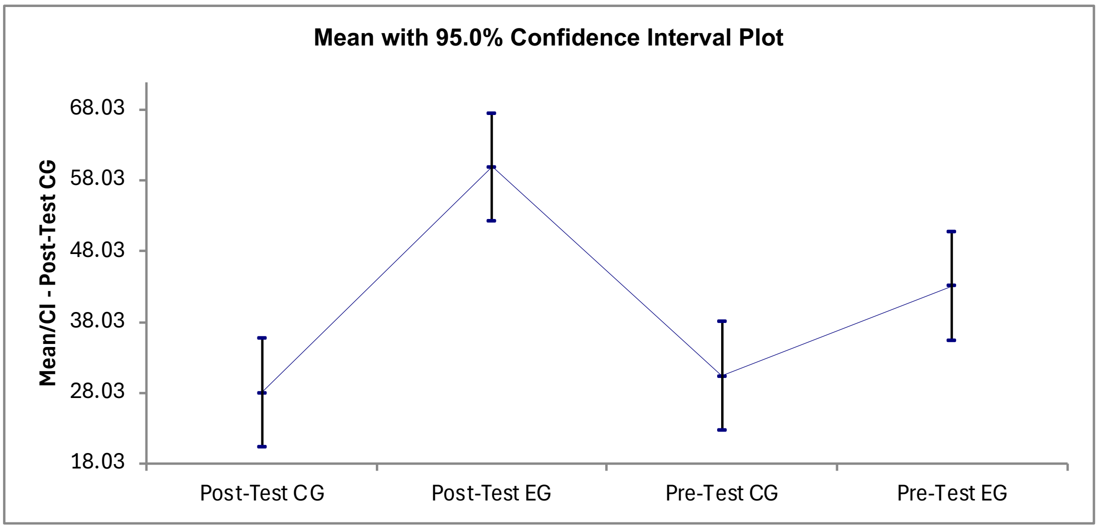
Table 16. ANOVA results for grade 3 MSAP jump strategy

Source	SS	DF	MS	F	P-value
Between	686.7767	3.0000	228.9256	3.238	0.0500
Within	1131.1784	16.0000	70.6987		
Total	1817.9551	19.0000			

The results indicate that both the control group and experimental group improved from pre-test to post-test (see Figure 21). The ANOVA results suggest a potential difference among the group means, but the significance is marginal ($p = 0.0500$). This borderline p-value indicates that the differences observed may not be robust enough to be considered strongly statistically significant.

Grade 3 MSAP bridge strategy

Figure 22. Trends in performance from pre to post-test between **the experimental group and control group for grade 3 MSAP bridge strategy**



The results indicate positive shifts in performance for the experimental group from pre-test to post-test compared to the control group (see Figure 22). The experimental group, influenced by the intervention, outperformed the control group.

Table 17. ANOVA results for grade 3 MSAP bridge strategy

Source	SS	DF	MS	F	P-value
Between	3203.5246	3.0000	1067.8415	16.390	**0.0000**
Within	1042.4245	16.0000	65.1515		
Total	4245.9491	19.0000			

The ANOVA results demonstrate a statistically significant difference among the group means ($p < 0.001$) (see Table 17). The high F-value (16.390) and extremely low p-value indicate that the observed differences in means between the groups are substantial and unlikely to be due to random variation. This result suggests that the group means differ significantly, and any observed differences between the group means are meaningful and can be attributed to the intervention applied. These results provide strong evidence of a statistically significant effect of the intervention.

9.2 Implementation Fidelity and Its Impact on Performance

Carroll et al. (2007) argue that the issue of implementation fidelity is mainly that of adherence to the intention of the intervention designers during implementation. They note that the key elements for the measurement of implementation fidelity consist of content, frequency, duration, and coverage. In our project on mental mathematics, we conceived of implementation fidelity as the degree of adherence to three things:

1) The teacher notes that detail how each session (tasks) should be taught following our three-step model (warm-up; whole class; individual/group activities)
2) The mathematics examples (questions) that are provided in the teacher and student books
3) The structure of the mathematics examples that are provided in the teacher and student books (for example, in a bridging through ten task, a teacher who uses examples such as 8 + 7 =__; 15 + 6 =__; 12 + 9 =__ (instead of 7 + 5 =__; 14 + 8 =__; 16 + 5 =__, which are in the teacher and student books) has maintained the structure of the task)

In our analysis of data, fidelity to structure emerged as the most important determinant of implementation fidelity much more than strict adherence to using the examples (questions) as they are in the teacher and student books. Through inductive and deductive processes, four categories of implementation fidelity emerged from our study: high fidelity, moderate fidelity, low fidelity, and no fidelity. Below, we present the descriptors and instances for some of these categories.

Table 18. Fidelity categories, descriptors, and instances

Category	Descriptors	Instances
High fidelity	• Exact examples (questions) from materials are used • Three-step model is used correctly • There is alignment between all used warm-up examples and lesson examples	I(i): 43 + 25 = EX:(w): 36 + 7 =
Moderate fidelity	• Similar structure of mathematics examples to implementation materials is used in examples given to the students (that is, structure is similar, questions (digits used) are different) • Most warm-up examples are aligned to lesson examples • Three-step model is used correctly	B(w): 35 – 28 = EM(p): 43 + 14 =
Low fidelity	• Different structure from the mathematics examples/questions is used • Similar digits are used (but with different structure from the mathematics examples in the teacher and student books) • Few examples in warm-up are aligned to examples in the main mental math lesson	
No fidelity	• Examples used are different from the ones in the teacher and student books • The examples used do not have a similar structure to the ones in teacher and student books • Three-step model is not used or used incorrectly • No warm-up exercises	

Tables 19 and 20 provide a comprehensive analysis of implementation fidelity, specifically examining tasks and instructional strategies in mental mathematics (see Appendix 4 for detailed analysis). The findings indicate that teachers generally implemented the materials with either high or moderate fidelity by adhering to the intended instructional activities outlined in the student and teacher books. For instance, two teachers consistently demonstrated high fidelity across the three days of observation, maintaining alignment with the designated tasks and problem-solving approaches, which included both individual and group/pair work. Additionally, one teacher exhibited high fidelity in most activities, particularly in applying the jump strategy and doubling and halving. While some teachers displayed strong fidelity in adhering to high-fidelity tasks, certain deviations occurred, leading to moderate fidelity in some instances. For example, in *School B* the teacher primarily emphasized rounding and adjusting strategies, maintaining a moderate fidelity level throughout. Similarly, in *School E-M*, the teacher also displayed moderate fidelity, reflecting partial adherence to the intended instructional approach.

Table 19. Implementation fidelity with three-step model: BTT

		Tasks and fidelity		
School name	**Day and strategy**	**Warm-up**	**Whole class**	**Individual/pair/group**
I	Day 1: Add 20s and 1s	No fidelity	High fidelity	High fidelity
	Day 2: Subtract 20s and 1s	Moderate fidelity	High fidelity	High fidelity
	Day 3: BTT	No fidelity	High fidelity	High fidelity

Table 20. Implementation fidelity with three-step model: MSAP

		Tasks and fidelity		
School name	**Day and strategy**	**Warm-up**	**Whole class**	**Individual/pair/group**
B	Day 1: Rounding and adjusting	No fidelity	Moderate Fidelity	Moderate Fidelity
	Day 2: Rounding and adjusting	No fidelity	Moderate fidelity	
	Day 3: Rounding and adjusting	No fidelity	Moderate fidelity	Moderate fidelity
E-M	Day 1: Rounding and adjusting	Moderate fidelity	Moderate fidelity	Moderate fidelity
	Day 2: BTT	Moderate fidelity	Moderate fidelity	Moderate fidelity
	Day 3: Jump strategy	Moderate fidelity	Moderate fidelity	Moderate fidelity
E-X	Day 1: BTT	High fidelity	High fidelity	
	Day 2: Jump strategy Rounding and adjusting	High fidelity	High fidelity	High fidelity
	Day 3: Doubling and halving Linking addition and subtraction Doubling and halving	High fidelity	High fidelity	Moderate Fidelity
N	Day 1: Jump strategy	High fidelity	High fidelity	High fidelity
	Day 2: Rounding and adjusting	High fidelity	High fidelity	Moderate fidelity
	Day 3: Doubling and halving	High fidelity	High fidelity	Moderate fidelity

Table 21 presents a summary of pre- and post-test mean scores across five schools, evaluating the implementation fidelity of mental mathematics tasks. These schools are categorized by fidelity levels—moderate or high—to instructional strategies in mental mathematics, and scores are compared to assess the impact on student performance.

For *School N*, 55 students took the pre- and post-tests on jump strategy. Overall, the teacher showed high implementation fidelity, and students had copies of the materials readily printed, which they could write on. The teacher had all the questions to be used (with other instructions) written on the board, prior to the lesson each day. There was evidence of alignment between the warm-up and the tasks throughout. On each day, the teacher worked with the students through the warm-up, followed by whole-class activities, group work, and pair work. The results for this teacher show a statistically significant mean score of 16.242 (p = 0.0000). The low p-value indicates a strong likelihood that this observed mean is significantly different from the hypothesized mean. The standard deviation ranges also indicate substantial increases, from -8.8 to 76.9 in the pre-test and 4.9 to 95.7 in the post-test. For *School I*, with moderate fidelity to the grade 2 BTT, there is a substantial mean improvement of 45.1 from pre- to post-test (mean pre: 21.7; mean post: 66.8). The standard deviation analysis shows a pre-test range extending from 1.2 to 42.2 and a post-test range from 33.6 to 100.0, indicating a broad increase in score distribution.

School B, also demonstrating moderate fidelity, shows mixed outcomes. Thirty students in this school took both the pre- and post-test on bridging through ten and jump strategies. This is the only school that submitted both strategies that the project focuses on. Overall, in terms of implementation fidelity, the teacher showed moderate fidelity. Despite not using any warm-up across the three days, the teacher, however, used questions that had a similar structure to those used in the materials. In BTT, the mean post-test score decreased by 13.9 (mean pre: 61.0; mean post: 47.1). In contrast, jump strategy tasks show a mean improvement of 13.8, rising from 13.9 to 27.7, though both tests display significant variance.

For both the pre- and post-tests, the p-value of 0.0026 is less than 0.05, meaning that the null hypothesis is rejected and a lot of the scores in the dataset are clustered around the mean. This indicates a statistically significant difference between the pre-test and post-test scores. Given this statistical analysis, it can be concluded that the teacher's ability to use questions that are similar in structure to those used in the materials, without the use of the warm-up, has somewhat led to a decline in the one strategy and an increase in the other. This perhaps suggests that the warm-up is an important part of the three-step model; in fact, all parts of the three-step model, when used correctly, could have led to better results.

Table 21. Summary of pre- and post-test mean scores

School		Mean-pre	Mean-post	Mean difference		Std. dev	Mean -2 std. dev	Mean +2 std. dev
I (15) High fidelity	BTT (grade 2)	21.7	66.8	45.1	**Pre**	10.3	1.2	42.2
					Post	16.6	33.6	100.0
B (30) Moderate fidelity	BTT	61.0	47.1	-13.9	**Pre**	21.6	18.4	103.6
					Post	23.5	0.9	93.3
	JS	13.9	27.7	13.8	**Pre**	15.5	-16.5	44.3
					Post	25.4	-27.5	82.8
E-M (27) Moderate fidelity	BTT	[redacted]						
	JS	19.6	30.2	10.6	**Pre**	16.4	-12.6	53.0
					Post	21.6	-14.6	71.8
E-X (28) High fidelity	BTT	[redacted]						
	JS	20.2	28.6	8.4	**Pre**	18.3	-17.0	56.3
					Post	25.3	-20.4	80.8
N (55) High fidelity	BTT	[redacted]						
	JS	34.1	50.3	16.2	**Pre**	21.4	-8.8	76.9
					Post	22.7	4.9	95.7

Table 22. Paired t-test results: Pre- and post-test

	I (JS)	B		E-M (JS)	E-X (JS)	N
		BTT	JS			
Count	15	30	30	27	28	55
Mean	45.067	-13.889	13.778	10.617	8.333	16.242
Std. dev	14.518	23.112	30.647	12.610	16.114	22.744
SE mean	3.749	4.220	5.595	2.427	3.045	3.067
t	12.022	-3.291	2.462	4.375	2.736	5.296
P-value (2-sided)	**0.0000**	**0.0026**	**0.0200**	**0.0002**	**0.0108**	**0.0000**

For *School E*, the results vary slightly between the two classes. Teacher E-M shows a moderate fidelity level in jump strategy tasks, with a mean gain of 10.6 (mean pre: 19.6; mean post: 30.2), with pre- and post-test ranges indicating modest score increases. Meanwhile, teacher E-X, marked as high fidelity, also shows gains, with a mean improvement of 8.4 in jump strategy tasks (mean pre: 20.2; mean post: 28.6). In teacher E-M's class, 28 students took both the pre- and post-tests on the jump strategy. Overall, the teacher used the three-step model correctly. The questions that the teacher mainly used were those that students had to give to the teacher. Generally, the teacher would give students the first question only for each part of mental math on each day (of the three days); thereafter, the questions that were used were those that students made up (all the questions had a similar structure to those in the materials).

In one instance, during the warm-up, one student gave a question that was exactly the same as one in the materials, which we found to be quite interesting. This is because by focusing on only this instance, our descriptors would point to this as high fidelity. However, given that in this particular class, students were not using copies of the materials (student book), it is interesting that this would now be classified as moderate fidelity. Had the question/example been given by the teacher, we would have classified this as high fidelity because the teacher had access to the materials. Furthermore, this particular teacher showed a deeper understanding of what the structure of questions that each strategy required, and through this, the teacher was able to tell/correct/restate the type of question (in terms of structure) she wanted students to give. The teacher rejected any questions that did not have or follow the particular structure of her planned lesson by reminding students of the conditions of what is required. Students' performance improved from pre- to post-test. The p-value of 0.0002 is much less than 0.05, which means that the null hypothesis (that there is no difference between post-test and pre-test scores) is rejected. There is a statistically significant difference between the two tests. The mean difference of 10.617 suggests that, on average, the scores improved from the pre-test to the post-test. In conclusion, the paired t-test results indicate a significant improvement in scores from the pre-test to the post-test. The positive mean difference and the significant p-value both support the conclusion that the intervention had a meaningful impact.

In teacher E-X's class, 27 students wrote both pre-and post-tests for the jump strategy. Overall, in terms of implementation fidelity, the teacher showed high fidelity throughout, except for one instance with individual tasks, where she showed moderate fidelity. Students' performance increased from pre- to post-test. For this teacher, the p-value of 0.0108 is less than 0.05, which means that the null hypothesis (that there is no difference between post-test and pre-test scores) is rejected. The mean difference of 8.333 suggests that, on average, scores increased from the pre-test to the post-test. The 95% confidence interval (2.085 to 14.582) means that we are 95% confident that the true mean difference lies within this range, and it does not include 0, further supporting the finding that there is a statistically significant difference between the pre-test and post-test scores. In conclusion, the paired t-test results suggest a significant improvement in scores from the pre- to post-test.

School I had 15 students who took both the pre- and post-tests (this is the only grade 2 school). Overall, in terms of implementation fidelity, the teacher showed high fidelity across the three-step model. In terms of the warm-up, the teacher did not use the exact questions entirely. In some instances, the teacher showed no fidelity during the warm-up, especially in instances where the warm-up was not aligned with the task itself. In terms of whole-class and individual activities, the teacher showed high fidelity across the three days. This means that for all the written work, the teacher gave students questions from tasks exactly as they are in the materials to solve. The students' performance increased from pre- to post-test, and the biggest increase from pre- to post-test is in this school. The results for this teacher's class reveal a statistically significant mean score of 45.067 ($p = 0.0000$), and since the p-value is less than 0.05, the null hypothesis can be rejected, and the results are statistically significant. Therefore, given the teacher's implementation fidelity and statistical data, it can be concluded that the use of exact examples has supported students' understanding of the mathematical concepts taught—hence the improvement in performance.

In summary, contrary to findings by Essien, Venkat, et al. (2015) (where teachers with moderate implementation fidelity tended to be more successful in teaching mathematical concepts due to teacher agency), higher-fidelity schools in our study, such as *School N* and *School E-X*, tended to demonstrate stronger gains in student performance, particularly in jump strategy tasks, compared to schools with moderate fidelity. This pattern suggests a potential correlation between implementation fidelity and student achievement in mental mathematics tasks.

10. Conclusion

The findings of this study highlight the impact of BTT strategies in improving students' performance in mental mathematics. The intervention led to statistically significant improvements in student performance, particularly in experimental groups compared to control groups, as evidenced by trends from pre- to post-tests, and statistically significant results from the ANOVA analysis. These findings suggest that the observed improvements are unlikely to be due to chance reinforcing the robustness of the intervention. Implementation fidelity emerged as a critical factor in the success of the intervention. Schools' adhering closely to the intervention model such as *School N* and *School E-X*, demonstrated marked improvements in jump strategy tasks. The structured use of intervention materials, particularly alignment with the three-step instructional model (warm-up, whole-class activities, and individual/group tasks), was consistently associated with substantial performance gains. Schools like *School B* and *School E-M* with moderate fidelity, showed mixed results. While some gains were observed, deviations from the intended instructional approach, such as the absence of warm-up activities or the use of similar but not identical question structures, appeared to limit the effectiveness of the intervention. The findings suggest that full adherence to all elements of the three-step intervention model, including the warm-up component, is essential for maximizing student performance.

Statistical analyses further validated these findings. For instance, paired t-test scores across schools revealed significant improvements in post-test scores in high-fidelity contexts, with p-values consistently below 0.05. For example, *School I* demonstrated the highest improvement, underscoring the efficacy of using exact examples aligned with intervention materials. Conversely, moderate fidelity was associated with mixed performance gains, highlighting the potential for further refinement in implementation practices. Overall, these findings underscore the importance of implementation fidelity in the success of educational interventions, particularly those (like ours) that have gone through different stages of iteration through design research. The results suggest that adhering to the designed instructional strategies, particularly the structured use of tasks and alignment across the three-step model, is crucial for fostering meaningful improvements in student performance. Moreover, the integration of BTT strategies into classroom practice not only enhances students' conceptual understanding of base-ten thinking but also fosters collaborative and active learning environments. To sustain these benefits, future iterations of similar interventions should priorities a comprehensive teacher training and support to maintain and ensure consistency in implementation fidelity. Also, an ongoing support mechanism that reinforces adherence to the instructional model should be considered. The BTT intervention provides a robust framework for improving foundational numeracy while emphasizing the role of well-structured instructional strategies. This approach not only drives academic success but also equips students with essential skills for lifelong learning, underscoring the broader impact of high-fidelity implementation in educational practices. In conclusion, the study highlights the transformative potential of high-fidelity implementation in educational interventions. By prioritizing structured, research-informed instructional strategies, interventions like this one can drive academic success and address broader challenges in mathematics education, ultimately supporting equitable and sustainable outcomes.

11. References

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher*, *41*(1), 16–25.

Anghileri, J. (2006). *Principles and practice in arithmetic teaching*. London: Routledge.

Askew, M., Graven, M., & Venkat, H. (2022). From what works to scaling up: Improving mental strategies in South African grade 3 classes.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, *2*, 1-9.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Sage Publications.

Essien, A. A., Mathews, C., Tshesane, H., Weitz, M., Abdulhamid, L., Hoosen, T., & Lavans, L. (2023). Developing design principles to enhance pre-service teachers' understanding of number structure and mathematical equivalence in early grade mathematics. *African Journal of Research in Mathematics, Science and Technology Education*, *27*(3), 239–254. 10.1080/18117295.2023.2216608

Essien, A. A., Venkat, H., Takane, T., & Tshesane, H. (2015). *An evaluation of the use and efficacy of the GPLMS FP multilingual mathematics materials: Issues and prospects—Final report*. Tshikululu.

Fritz, A., Long, C., Herzog, M., Balzer, L., Ehlert, A., & Henning, E. (2020). Mismatch of the South African Foundation Phase curriculum demands and learners' current knowledge. *African Journal of Research in Mathematics, Science and Technology Education*, *24*(1), 10–20. 10.1080/18117295.2020.1724466

Fullan, M. G. (1993). Why teachers must become change agents. *Educational Leadership*, *50*(6), 1–13.

Graven, M., & Venkat, H. (2021). Piloting national diagnostic assessment for strategic calculation. *Mathematics Education Research Journal*, *33*(1), 23–42.

Hartley, Z. (2007). *Setting a strong foundation in literacy and numeracy up to grade 6 through a comprehensive GET strategy*. Western Cape Education Department

Hoadley, U. (2007). The reproduction of social class inequalities through mathematics pedagogies in South African primary schools. *Journal of Curriculum Studies*, *39*(6), 679–706. https://doi.org/10.1080/002202707012611694

Human, A., Van der Walt, M., & Posthuma, B. (2015). International comparisons of Foundation Phase number domain mathematics knowledge and practice standards. *South African Journal of Education*, *35*(1), 1–13.

Johnson, R. B., & Christensen, L. (2014). *Educational research: Quantitative, qualitative, and mixed approaches* (5th ed.). Sage Publications.

Knuth, E., A. Stephens, A., N. McNeil, N., & Alibali, M. (2006). Does understanding the equal sign matter? Evidence from solving equations. *Journal for Research in Mathematics Education*, *37*(4), 297–312.

Mason, J., Stephens, M. & Watson, A. (2009). Appreciating mathematical structure for all. *Mathematics Education Research Journal 2009*, *21*(2), 10–32.

Mohohlwane, N., & Taylor, S. (2015). Using impact evaluation for education policy innovations: The case of early grade literacy in South Africa. eVALUatiOn Matters.

Morrison, S. (2018). Exploring the role of base ten thinking in learning mathematics: A study with grade 2 learners. Doctoral dissertation.

Mulligan, J. (2002). The role of structure in children's development of multiplicative reasoning. In B. Barton, K. C. Irwin, M. Pfannkuch, & M. O. Thomas (Eds.), *Mathematics education in the South Pacific: Proceedings of the 25th annual conference of the Mathematics Education Research Group of Australasia* (Vol. 2, pp. 497–503). MERGA.

Mulligan, J., & Mitchelmore, M. (2009). Awareness of pattern and structure in early mathematical development. *Mathematics Education Research Journal*, *21*(2), 33–49.

Schollar, E. (2008). *Teaching and learning mathematics in South Africa: The impact of curriculum reform*. Human Sciences Research Council Press.

Spaull, N., & Kotze, J. (2015). Starting behind and staying behind in South Africa: The case of insurmountable learning deficits in mathematics. *International Journal of Educational Development*, *41*, 13–24. http://doi.org/10.1016/j.ijedudev.2015.01.002

Weitz, M., & Venkat, H. (2013). Assessing early number learning: How useful is the Annual National Assessment in Numeracy? *Perspectives in Education*, *31*(3), 49–65.

Wright, R. J., Martland, J., & Stafford, A. (2012). *Teaching number: Advancing children's skills and strategies*. Sage Publications.

Appendix 1: Test Items – Jump Strategy

Name:

Jump Strategies: Pre-Test

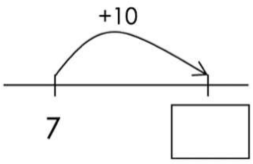
PART I — 2 minutes for this page

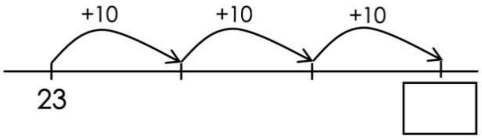
1. Fill in the missing number
14, 24, 34, 44, ☐

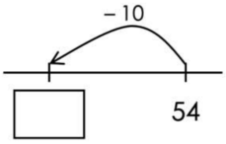
2. Fill in the missing number
79, 69, 59, 49, ☐

3. 6 + 30 = ☐

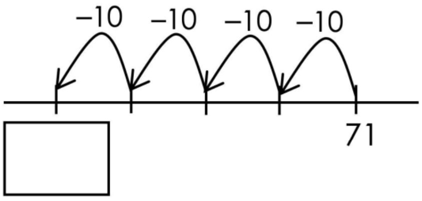
4. 57 – 10 = ☐

5. 

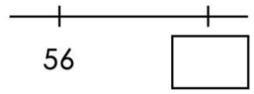
6. 

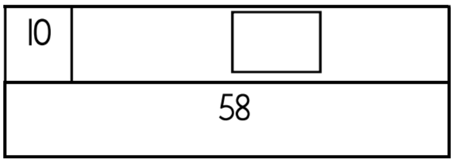
7. 

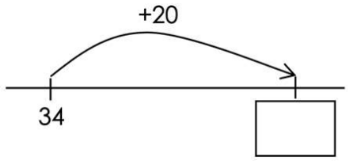
8. 36 + ☐ = 40

9. 

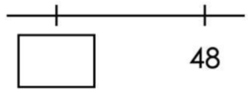
10. 31 – 20 = ☐

11. What is the next multiple of 10?


12. 

13. 

14. 16 + 30 = ☐

15. What is the multiple of 10 before 48?


16. 79 – 40 = ☐

17. 38 – ☐ = 18

18. ☐ – 20 = 69

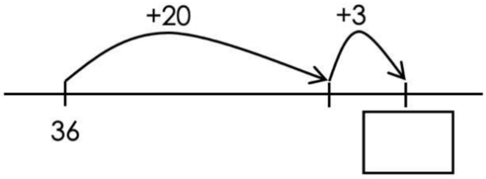
19. 37 + ☐ = 77

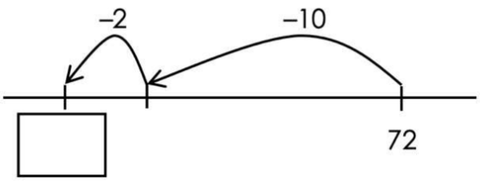
20. ☐ + 20 = 66

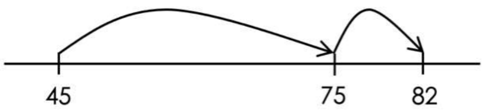
Total out of 20

Jump Strategies: Pre-Test
PART 2

3 minutes for this page

1.


2.


3.


45 + ☐ + 7 = 82

4.


53 – ☐ – 4 = 29

5. 57 + 26 = ☐

6. 83 – 24 = ☐

7. 19 + ☐ = 41

8. 62 – ☐ = 47

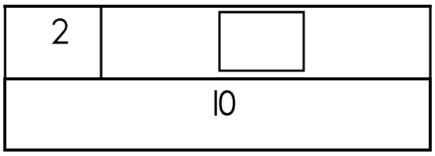
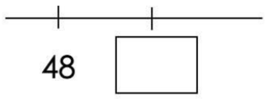
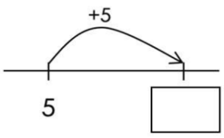
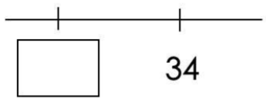
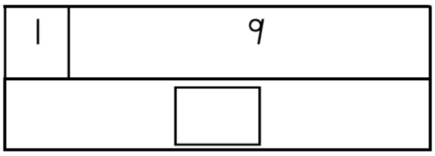
9. 61 – 32 = 61 – ☐ – 2

10. 74 – ☐ = 74 – 20 – 5

Total out of 10	

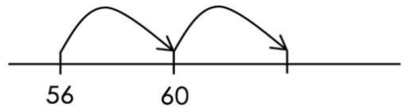
Appendix 2: Test Items – Bridge Strategy

Bridging Through Ten

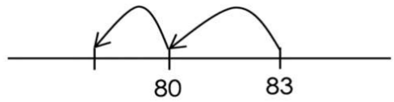
Name:			
Bridging Through Ten: Pre-Test PART 1			2 minutes for this page
1.	7 + 3 = ☐	11.	50 + 6 = ☐
2.	2 + 8 = ☐	12.	3 + 60 = ☐
3.	10 = 7 + ☐	13.	40 – 7 = ☐
4.	8 less than 10 is ☐	14.	40 + 8 = ☐
5.		15.	What is the next multiple of 10? 
6.		16.	100 + 27 = ☐
7.	10 – 5 = ☐	17.	What is the multiple of 10 before 34? 
8.	10 – 4 = ☐	18.	☐ + 7 = 50
9.		19.	30 – ☐ = 27
10.	☐ + 10 = 10	20.	87 = 80 + ☐
Total out of 20			

Bridging Through Ten: Pre-Test
PART 2 — 3 minutes for this page

1. 56 + 8 = ☐



2. 83 – 4 = ☐



3. 93 – 7 = ☐

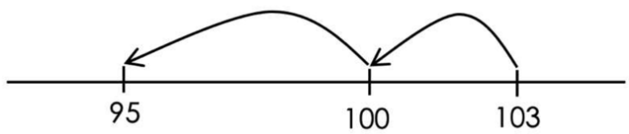
4. 67 + ☐ = 73

5. ☐ + 7 = 82

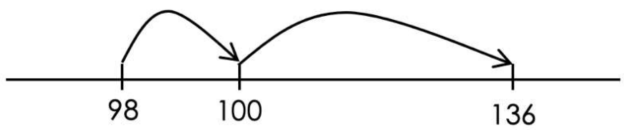
6. 67 + 5 = 67 + 3 + ☐

7. 94 – ☐ = 94 – 4 – 2

8. 98 + 56 = 98 + 2 + ☐

9. 

103 – ☐ = 95

10. 

98 + ☐ = 136

Total out of 10	

Appendix 3: Test Items – BTT

* Questions 1–5 not included because they are not in the public domain.

Q6 - 2 minutes

6a	9 + 10 =		6e	9 + 100 =	
6b	92 + 10 =		6f	42 + 100 =	
6c	50 – 10 =		6g	120 – 100 =	
6d	700 – 10 =		6h	400 – 100 =	

Q7 - 1 minute

7a	30 + 20 = 50	7b	30 + 20 = 50	7c	30 + 20 = 50
	30 + 21 = ____		32 + 20 = ____		30 + 19 = ____

Q8 - 1 minute

8a	34 – 10 = 24	8b	34 – 10 = 24	8c	34 – 10 = 24
	34 – 20 = ____		34 – 12 = ____		34 – 15 = ____

Q9 - 1 minute

						A			
			14						
			B						
	C								

A = ____

B = ____

C = ____

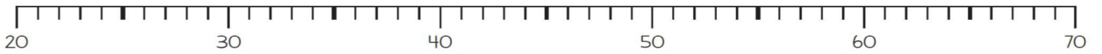
B + 20 = ____

B + 12 = ____

C – 10 = ____

Q10 - 3 minutes

10a 37 + 21 = ____

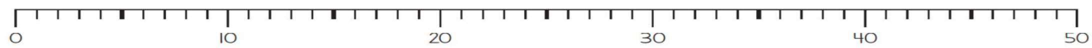


10b 7 + ____ = 41

10c 36 + 25 = ____

10d ____ + 12 = 39

10e 35 – 21 = ____



10f 49 – ____ = 34

Appendix 4: Implementation Fidelity

BTT: Table of implementation fidelity with examples

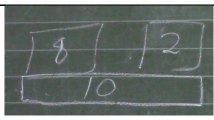
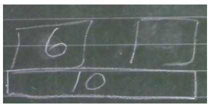
		Tasks and fidelity		
School name	**Day and strategy**	**Warm-up**	**Whole class**	**Individual/pair/group**
I	Day 1: Add 20s and 1s	Teacher asked students to give a combination of two numbers that make a 10, 20, and 15. 10: 7 + 3 = 5 + 5 = 9 + 1 = 8 + 2 = 20: 10 + 1 0= 17 + 3 = 15 + 5 = 19 + 1 = 15: 14 + 1 = 10 + 5 = 13 + 2 = 12 + 3 = No fidelity	Task 60 – p. 47 34 + 24 = 51 + 23 = High fidelity	Individual Task 60 – p.47 43 + 2 5= 32 + 24 = High fidelity
	Day 2: Subtract 20s and 1s	Teacher asked students to give a combination of two numbers that make a 30, 15, and 12. 30: 20 + 10 = 29 + 1 = 25 + 5 = 30 + 0 = 28 + 2 =	Task 64 – p. 50 47 - 23 = 58 - 23 = High fidelity	Individual Task 64 – p. 50 46 - 21 = 57 - 24 = High fidelity

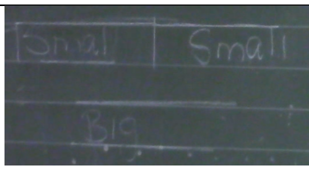
		15: 10 + 5 = 14 + 1 = 13 + 2 = 6 + 9 = 7 + 8 = 12: 11 + 1 = 10 + 2 = 8 + 4 = 9 + 3 = 6 + 6 = 12 + 0 = Subtract 20 from a given number: 100 - 20 = 130 - 20 = 90 - 20 = 74 - 20 = 54 - 20 = Moderate fidelity		
	Day 3: BTT	Teacher asked students to give a combination of two numbers that make a 10. 10: 7 + 3 = 9 + 1 = 8 + 2 = 5 + 5 = 6 + 4 = 1 + 9 =	Task 73 – p. 56 26 + 9 = 28 + 6 = High fidelity	Individual Task 73 – p. 56 37 + 8 = 45 + 7 = High fidelity

		10 + 0 = No fidelity		

MSAP: Table of implementation fidelity with examples

		Tasks and fidelity		
School name	**Day and strategy**	**Warm-up**	**Whole class**	**Individual/pair/group**
B	Day 1: Rounding and adjusting	No fidelity	35 - 28 = 39 - 25 = Moderate fidelity	Individual work 47 - 29 = (pp. 82, 91) 37 - 28 = Moderate fidelity
	Day 2: Rounding and adjusting	No fidelity	28 + 28 + 39 = 37 + 49 + 27 = Moderate fidelity	Individual work 26 + 37 + 49 = 27 + 68 + 19 = 28 + 39 + 57 =
	Day 3: Rounding and adjusting	No fidelity	56 + 75 = 47 + 58 + 59 = 55 + 46 + 57 = 67 + 15 + 38 = Moderate fidelity	Individual work 47 + 58 + 39 = Moderate fidelity
E-M	Day 1: Rounding and adjusting	Round off the given number to the nearest 10. 27→30 (by teacher) 39→40 (by teacher) 57→60 (by students) 77→80 (by students) 37→40 (by students) 87→ 90 (by students) Moderate fidelity	19 + 39 = (by teacher) 48 + 18 = (by students) 87 + 38 = (by students) 57 + 16 = (by students) Moderate fidelity	18 + 18 + 18 = (by teacher) 36 + 36 + 36 = (by students) 49 + 49 + 49 = (by students) 57 + 57 + 57 = (by students) Moderate fidelity

	Day 2: BTT	Round off the given number to the nearest 10. 21→30 42→50 61→70 83→90 45→50 Moderate fidelity	36 + 7 = 95 + 9 = Moderate fidelity	<u>Pair work – given by students</u> 84 + 8 = 78 + 9 = 67 + 8 = Moderate fidelity
	Day 3: Jump strategy	23 + 10 = 44 + 10 = 11→21 12→22 82→92 78→88 32→42 Moderate fidelity	38 + 9 = 36 + 13 = 48 + 16 = (by students) 64 + 18 = (by students) Moderate fidelity	<u>Pair work</u> 43 + 14 = 73 + 17 = 67 + 23 = 45 + 39 = Moderate fidelity
E-X	Day 1: BTT	  (p. 2) High fidelity	36 + 7 = (p. 7) 35 + 8 =(p. 9) 27 + 8 =(p. 7) High fidelity	
	Day 2: Jump strategy Rounding and adjusting	43 - 12 = (p. 41) High fidelity	36 - 7 = (p. 19) 35 - 8 = (p. 19) 27 + 9 = (g. 80) 34 + 9 = (p. 80) High fidelity	Individual work 62 + 9 = (p. 80) High fidelity
	Day 3: Doubling and halving	Double: e.g., 10 (p. 52)	Doubling and halving 6 + 6 = (p. 52) 8 + 8 = (p. 52)	Individual work Double: 42 (p. 53); 35 (p. 61); 26

	Linking addition and subtraction Doubling and halving	 (p. 130) 6 + … = 10 6 - 3 = (p. 131) 9 = … + 6 6 = 3 = 9 – 6 = (p. 131) 4 + 6 = (p. 132) 10 - 6 = (p. 132) High fidelity	Half of 10 (p. 57) Half of 6 50 + 50 = 20 + 20 = Double: 22; 36; 47 (p. 62) Double: 29 (p. 62); 34 High fidelity	Linking addition and subtraction 4 + 6 = (p. 132) …+ 4 = 10 10 - …= 6 …+ 6 = 9 6 -…= 9 Moderate fidelity
N	Day 1: Jump strategy	-Give the next multiple of 10. -Give a number to be added to give a multiple of 10. High fidelity	47 + 29 = (p. 82) Adjust 47 + 29 = High fidelity	Individual work 54 + 39 = (p. 83) 26 + 19 = (p. 83) 74 + 36 = High fidelity
	Day 2: Rounding and adjusting	Count on in 10s. High fidelity	19 + 39 = (p. 94) 28 + 49 = (p. 94) 99 + 99 = (p. 98) High fidelity	Group work 38 + 39 = 19 + 19 +19= (p. 92) Individual work 68 + 27 = Moderate fidelity
	Day 3: Doubling and halving	Doubles up to double: e.g., 3; 4; 5; 6 Half of: e.g., 8 (p. 56) High fidelity	Double: 35 (p. 61); 29 (p. 62) Half of: 62 (p. 66); 76 (p. 66) High fidelity	Group work Double: 41; 36 Half of: 68 (p. 66); 82 Individual work Double: 27; 43 Half of: 26; 88 Moderate fidelity