MODESTUM EURASIA Journal of Mathematics, Science and Technology Education, 2026, 22(1), em2767
ISSN:1305-8223 (online)
OPEN ACCESS Research Paper https://doi.org/10.29333/ejmste/17759

Student readiness for TIMSS through the lens of mathematical proficiency: Evaluation of an instructional initiative for grades 5-8 in the UAE

Rahila Naseeha Abul Kalaam [1], Jogymol Kalariparampil Alex [1*]

[1] United Arab Emirates University, Al Ain, UAE

Received 13 October 2025 ▪ Accepted 08 December 2025

Abstract

This study investigated a targeted educational intervention in the United Arab Emirates (UAE) aimed at enhancing students' mathematical proficiency and readiness for the trends in international mathematics and science study (TIMSS). Grounded in Kilpatrick's five interwoven strands of mathematical proficiency–conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition–the research employed a quantitative one-group pre-test/post-test design to evaluate learning gains. Despite ongoing national reforms, UAE students continue to demonstrate challenges in higher-order mathematical thinking, highlighting a need for targeted, evidence-based interventions. Data were collected from low-achieving students across grade 5-grade 8 to measure strand-based improvements and their alignment with TIMSS cognitive domains. Results revealed that students' mean mathematics scores improved by 8.7 points ($p < .001$), demonstrating measurable proficiency gains. These findings provide empirical evidence linking Kilpatrick's strands to quantifiable improvements in TIMSS-related skills within the UAE context and contribute to ongoing efforts to strengthen national mathematics education outcomes.

Keywords: TIMSS, mathematical proficiency, instructional initiative

INTRODUCTION

The escalating demands of the globalized economy and the imperative for an informed citizenry accentuate the critical role of robust mathematics education in fostering innovation and economic competitiveness. International assessments, such as the trends in international mathematics and science study (TIMSS), serve as vital benchmarks for evaluating the efficacy of educational systems worldwide, providing critical insights into student achievement and curriculum effectiveness (Pertiwi & Wahidin, 2020). Despite ongoing educational reforms and strategic initiatives in the United Arab Emirates (UAE) aimed at aligning educational outcomes with national aspirations for a knowledge-based economy (Jones, 2012), students consistently exhibit notable limitations in mathematical proficiency across various thematic areas. Specifically, UAE fourth and eighth graders have consistently scored below the TIMSS mathematics scale average (Balala et al., 2021), with low levels of mathematics proficiency evident among UAE learners (Almarashdi & Jarrah, 2023). Furthermore, UAE students demonstrate underperformance in international assessments related to problem-solving and critical thinking (Mizyed & Eccles, 2023), with a significant proportion scoring below basic proficiency levels. These challenges encompass foundational conceptual understanding in areas such as fractions, decimals, and percentages, where students frequently encounter difficulties (Singh et al., 2021), suggesting a pervasive deficit in solid mathematical comprehension. Addressing these deficiencies is paramount for the UAE to improve its standing in international assessments like TIMSS and to adequately equip its youth with essential mathematical skills for future challenges. International comparative studies highlight that student performance gaps in mathematics often stem not solely from curriculum breadth but critically from the depth of conceptual understanding and the sophistication of problem-solving skills cultivated in learners (Klang et al., 2021). Assessments reveal that a substantial portion of students struggle with problems requiring more than direct inference,

✉ 700039205@uaeu.ac.ae ✉ jalex@uaeu.ac.ae **(*Correspondence)**

Contribution to the literature

- This study contributes empirical evidence on the effectiveness of a strand-based instructional intervention grounded in Kilpatrick's model of mathematical proficiency for improving TIMSS-related performance among low-achieving students.
- It extends existing literature by quantitatively linking specific proficiency strands to measurable gains aligned with TIMSS cognitive domains, addressing a gap in intervention-focused research within middle-grade mathematics education.
- Situated in the UAE context, the findings offer a contextually grounded model for designing evidence-based instructional strategies to support higher-order mathematical thinking in reform-oriented education systems.

underscoring the importance of applied reasoning beyond rote knowledge (Klang et al., 2021). While instructional practices have traditionally prioritized procedural fluency–the ability to apply algorithms (Serhan, 2014)–a persistent lack of emphasis on fostering mathematical reasoning, conceptual clarity, and practical application continues to hinder overall student performance. Conceptual and procedural knowledge are integrally related, with a consensus that their relationship is bi-directional and iterative, both being essential for comprehensive mathematical proficiency (Rittle-Johnson & Schneider, 2014). Current educational reforms in the UAE, including enhanced teacher professional development and curriculum revisions, have been strategically implemented to mitigate these challenges. However, evidence consistently shows that UAE students display significantly lower performance on standardized mathematics assessments overall, indicating that low-achieving students remain particularly vulnerable in mathematics education, often necessitating additional support. Consequently, targeted pedagogical interventions that meticulously align with both Kilpatrick's interwoven strands of mathematical proficiency and the TIMSS cognitive domains are urgently needed. Such interventions are crucial for transcending a mere focus on procedural recall, thereby cultivating students' capacities for effective application and rigorous reasoning within diverse mathematical contexts. However, few empirical UAE studies have examined how short-term interventions influence TIMSS-related competencies across all five proficiency strands, leaving a critical evidence gap this study addresses. In this study, we conceptualize "TIMSS readiness" not as direct practice with released TIMSS items, but as strengthening the arithmetic strands of mathematical proficiency–particularly conceptual understanding, procedural fluency, and adaptive reasoning–that underpin performance on TIMSS number items and related applying and reasoning tasks.

Strands of Mathematical Proficiency

According to Kilpatrick et al. (2001), mathematical proficiency comprises five interrelated strands: conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition. Effective instructional design must simultaneously nurture conceptual understanding and procedural fluency while developing strategic competence, adaptive reasoning, and productive disposition (Adeniji & Baker, 2022; Miao et al., 2021). These strands are mutually reinforcing and form the foundation of meaningful mathematical learning. Conceptual understanding refers to comprehension of mathematical concepts, operations, and relations. In the UAE context, persistent difficulties in fractions, decimals, percentages, algebra, and geometry reflect limited conceptual depth among students (Mizyed & Eccles, 2023). Procedural fluency denotes the ability to carry out procedures accurately and efficiently. While many students demonstrate fluency in basic operations, their performance weakens in multi-step problems, indicating that fluency is not consistently transferable across domains. Strategic competence is the capacity to formulate, represent, and solve mathematical problems, an area where UAE students frequently struggle in applying strategies flexibly (Ibrahim & Alhosani, 2020). Adaptive reasoning involves logical thought, reflection, and justification, yet research indicates that students often have difficulty articulating or defending solutions–skills vital for higher-order mathematical reasoning (Yeliz, 2015). Finally, Productive Disposition refers to viewing mathematics as sensible and worthwhile, coupled with confidence in one's ability to succeed; repeated difficulties often reinforce negative attitudes that must be deliberately addressed (Moussa & Saali, 2022). These five strands together provide a robust framework for evaluating mathematical proficiency. Importantly, they align closely with the three TIMSS cognitive domains: *knowing* (linked to procedural fluency and basic conceptual understanding), *applying* (linked to strategic competence and adaptive reasoning), and *reasoning* (linked to advanced conceptual understanding and logical justification). This theoretical alignment guided both the construction of the study's mathematics proficiency test and the interpretation of its results. The test items were designed to reflect this interconnection–ensuring that each assessed strand corresponded to a TIMSS domain–thus enabling a coherent framework for analyzing improvements in mathematical proficiency and TIMSS readiness.

The UAE Intervention Context

In response to persistent weaknesses in mathematical proficiency, the UAE has initiated targeted educational interventions designed to strengthen students' foundational skills (Balala et al., 2021). These programs typically address critical gaps in areas such as fractions, decimals, percentages, and integer operations, while also extending to more complex algebraic and geometric reasoning (Wardat et al., 2022). Such interventions aim to not only foster procedural fluency–the ability to execute mathematical procedures accurately and efficiently (Adeniji & Baker, 2022)–but also to deepen conceptual understanding, recognizing that a balanced approach is essential for long-term improvement and effective problem-solving (Burns et al., 2015; Ho, 2020). Research indicates that both conceptual and procedural knowledge are integrally related and critical for comprehensive mathematical proficiency (Rittle-Johnson & Schneider, 2014). The implementation of web-based platforms (Shana et al., 2024) and AI-driven diagnostic tools (Miao et al., 2021) further supports these targeted approaches in UAE schools. Specifically, our intervention at a public school involved pre-service teachers delivering structured support to low-achieving students in grade 5-grade 8. Drawing on the framework of mathematical proficiency articulated by Kilpatrick et al. (2001), this program emphasized the interconnected strands of conceptual understanding, procedural fluency, strategic competence, and adaptive reasoning, as also noted in later studies like Altarawneh and Marei (2021) and Yulian and Wahyudin (2018). The analysis of this intervention offers valuable insights into the efficacy of structured approaches and their potential to inform future educational strategies within the UAE context. This is particularly relevant given prior evidence of positive impacts of interventions on low-achieving students (Betts et al., 2023; Subba & Gotamey, 2022). Moreover, these findings are consistent with TIMSS-derived evidence that reinforcing conceptual understanding and procedural fluency concurrently can significantly improve mathematics achievement among low-performing cohorts (Badri et al., 2020).

Problem Statement

Despite major curriculum reforms and investment in teacher professional development, persistent gaps remain in students' mathematics achievement across the UAE, particularly among low-achieving learners. International assessments such as TIMSS reveal that many UAE students struggle with foundational concepts and fail to demonstrate proficiency in higher-order reasoning and problem-solving. Since the UAE first participated in TIMSS in 2011, national average mathematics scores have improved but remain below the TIMSS scale average of 500. In TIMSS 2019, UAE students achieved average mathematics scores of 481 at grade 4 and 473 at grade 8, up from 434 and 456, respectively in 2011, yet still below the international center point of 500. In 2019, this placed the UAE 34th of 58 participating systems in grade 4 mathematics and 23rd of 39 in grade 8 mathematics, highlighting ongoing challenges in students' foundational understanding and reasoning (Mullis et al., 2020; UAE Ministry of Education, 2020). Prior studies have attributed these difficulties to limited opportunities for conceptual understanding, an overemphasis on procedural practice, and insufficient scaffolding for strategic and adaptive reasoning (Miao et al., 2021; Mizyed & Eccles, 2023). These findings emphasize the need for interventions that explicitly target all five strands of mathematical proficiency as identified by Kilpatrick et al. (2001). However, few empirical UAE-based studies have systematically examined how short-term instructional interventions influence TIMSS-related competencies across these strands. This study therefore addresses that critical gap by evaluating a six-week, strand-based mathematics intervention implemented among low-achieving students in grade 5-grade 8.

Research Questions

The study was guided by the following research questions on the strand–based intervention:

1. To what extent does the intervention improve students' overall mathematical proficiency and TIMSS readiness?
2. How does the intervention affect performance across Kilpatrick's strands of mathematical proficiency–particularly procedural fluency, conceptual understanding, strategic competence, and adaptive reasoning?
3. In what ways does the intervention foster productive disposition among low-achieving students toward mathematics learning?

MATERIALS AND METHODS

Research Design

This study was part of a bigger project on mathematics teaching and learning enhancement (MTLE) instructional initiative in a public school in the UAE. It adopted a quantitative one-group pre-test/post-test design to evaluate the effectiveness of a short-term mathematics intervention aimed at enhancing students' mathematical proficiency and readiness for the TIMSS. The mathematics proficiency test was therefore designed as a TIMSS-oriented, strand-based instrument: items targeted number and operations content typical of TIMSS grade 4-grade 8, and were mapped onto TIMSS cognitive domains (knowing, applying, and reasoning) while also being aligned to Kilpatrick's strands of

Table 1. Distribution of participants across grade levels

Grade level	G5	G6	G7	G8
Students (n)	28	18	15	12

proficiency. Rather than serving as a direct TIMSS practice test, the instrument operationalized the foundational competencies that contribute to TIMSS readiness. This design enabled direct comparison of performance before and after the intervention within the same cohort, providing practical evidence of learning gains in an authentic school context where randomization was not feasible. A control group was not included due to ethical and logistical constraints–specifically, the school's preference not to withhold potentially beneficial instruction from comparable students. To mitigate threats to internal validity, standardized testing conditions, equivalent pre- and post-test forms, and consistent instructional procedures were applied. The intervention's impact was evaluated across Kilpatrick et al.'s (2001) five strands of mathematical proficiency–conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition–aligned with the TIMSS cognitive domains of knowing, applying, and reasoning.

Participants

Participants were 73 low-achieving female students enrolled in grade 5 through grade 8 at a public girls' school in Al Ain, UAE. The participating school, located in a socio-educationally diverse area of Al Ain, serves predominantly Emirati students following the UAE Ministry of Education curriculum. Students were purposively selected based on teacher recommendations and consistent underperformance in mathematics assessments, representing learners most in need of academic support. **Table 1** presents the grade-wise distribution of participants.

Instruments

A mathematics proficiency test, developed by the researchers of the MTLE project served as both pre-test and post-test for this study. The test comprised 12 items (25 total points) spanning five arithmetic strands: multi-digit addition, subtraction, mental multiplication, distributive property application, and division with remainders. Each section corresponded to one or more of Kilpatrick's proficiency strands and TIMSS cognitive domains.

Content validity was established through expert review by three mathematics education specialists who verified the test's alignment with UAE learning outcomes and the TIMSS framework. Reliability was confirmed through a pilot administration to a comparable grade 5 cohort, yielding a Cronbach's alpha of 0.83, indicating strong internal consistency. To ensure scoring reliability, 25% of test papers were double-marked independently, achieving inter-rater agreement above 0.95.

Sample items included:

Procedural fluency: 6,369 + 45,245 = ?

Conceptual understanding: Explain why 12 × 14 = 12 × (10 + 4).

Adaptive reasoning: 374 ÷ 8 = ? Explain what the remainder means.

These sample items were designed to elicit different aspects of Kilpatrick's strands. The addition item requires students to carry out a multi-digit written algorithm accurately and efficiently, providing evidence of procedural fluency. The multiplication item asks students to explain why 12 × 14 can be expressed as 12 × (10 + 4) thereby assessing conceptual understanding of the distributive property and the relationship between symbolic expressions and their decompositions. The division item requires students not only to compute 374 ÷ 8 but also to interpret and explain the meaning of the remainder in context, tapping adaptive reasoning as students justify how their numerical result relates to a real-world situation.

In constructing these and the remaining items, we drew on the TIMSS assessment framework for grade 4 and grade 8. All test questions were aligned with the Number content domain and mapped to the three TIMSS cognitive domains–knowing, applying, and reasoning–with procedural fluency items primarily targeting knowing, items requiring explanation or representation targeting reasoning, and multi-step word problems targeting applying. In this way, the researcher-generated assessment did not replicate released TIMSS items, but it operationalized the same types of knowledge and processes that TIMSS seeks to measure.

Scoring and Strand Level Coding

Each of the 12 items was mapped to a priori to one primary proficiency strand (conceptual understanding, procedural fluency, strategic competence, or adaptive reasoning) and to a TIMSS cognitive domain (knowing, applying, or reasoning) using a test blueprint. Student responses were scored with an analytic rubric that allowed for partial credit. Depending on the complexity of the task, items carried 1-3 points, yielding a maximum total score of 25 points for the test. For example, straightforward computation items were scored 0-1 (incorrect/correct), whereas items requiring explanation or interpretation were scored 0-2 or 0-3, with higher scores reflecting both correctness and the quality of the student's reasoning or explanation.

For analysis, strand scores were obtained by summing the points from items mapped to each strand, resulting in separate pre- and post-test scores for conceptual understanding, procedural fluency, strategic competence, and adaptive reasoning. Because the test was intentionally brief, the strands were represented by different numbers of items; we therefore did not constrain each strand to have identical maximum scores. Instead, the blueprint ensured that all strands were assessed by multiple items and that the overall distribution of points reflected the centrality of whole-number operations in the target curriculum and in the TIMSS Number domain. Scoring did not reduce responses to a simple correct/incorrect dichotomy; rather, the analytic rubric captured the nature and quality of students' strategies, explanations, and interpretations, which allowed us to make strand-based inferences about changes in procedural fluency, conceptual understanding, strategic competence, and adaptive reasoning.

Data Collection Procedures

Data collection took place in three structured phases.

Phase I. Pre-test (baseline)

A 30-minute baseline test was administered under standardized conditions. Students completed the test independently without calculators, and all scripts were collected immediately after completion. To enhance reliability, a subsample was re-scored by a second rater, yielding nearly perfect agreement.

Phase II. Intervention

The six-week instructional intervention was delivered by trained pre-service mathematics teachers. The content focus was on core whole-number operations that had been identified as persistent weaknesses in school assessment reports and that are heavily represented in the TIMSS number domain, namely:

(a) multi-digit addition and subtraction with regrouping,

(b) mental and written multiplication with two- and three-digit whole numbers,

(c) application of the distributive property to expressions such as 8 × 53, and

(d) division with remainders, including interpretation of the remainder in word problems.

These topics mirrored the strands assessed in the mathematics proficiency test (multi-digit addition, subtraction, mental multiplication, distributive-property applications, and division with remainders).

Each 60-minute session was structured to target specific proficiency strands and to connect procedural fluency, conceptual understanding, and adaptive reasoning. Lessons typically began with a short conceptual warm-up (e.g., using number lines, area models, or part–whole diagrams) to surface students' prior ideas and make underlying concepts explicit. This was followed by focused practice of written algorithms to strengthen procedural fluency, accompanied by teacher prompts that asked students to explain *why* the procedures worked. In the main part of the lesson, students worked in pairs or small groups on strategy-based word problems that required them to select and justify operations, compare alternative solution paths, and interpret remainders. Whole-class discussions at the end of each session provided opportunities for students to articulate and defend their reasoning, thereby fostering adaptive reasoning.

Compared with conventional lessons in these classes, which teachers reported as being dominated by teacher demonstration followed by individual practice exercises, the intervention placed greater emphasis on

(a) explicit use of visual and conceptual models alongside algorithms,

(b) structured opportunities for students to explain and justify their thinking, and

(c) the use of formative assessment routines (warm-up questions, exit tickets, and quick-response tasks) to adjust pacing and revisit misconceptions.

To ensure fidelity of implementation, researchers conducted weekly observations using a structured observation guide and provided written and verbal feedback to the pre-service teachers.

Phase III. Post-test

At the end of the intervention, an equivalent post-test was administered using parallel forms. Administration conditions matched those of the pre-test. The same scoring rubric and reliability procedures were followed, and pre-/post-test results were paired through anonymized codes. All data were securely stored for analysis.

Data Analysis

Data analysis combined descriptive and inferential statistics. Individual pre- and post-test scores were computed both overall and by proficiency strand. Descriptive statistics (means, standard deviations, and gain scores) provided a summary of group performance. Paired-sample t-tests ($\alpha = .05$) were used to determine the significance of mean differences between pre- and post-test scores. Effect sizes (Cohen's d) were calculated to quantify the magnitude of observed improvements, with $d > 0.8$ interpreted as a large effect. Additionally, bar charts were used to visualize mean gains by strand and by grade level, enabling a comparative interpretation of performance trends.

Ethical Considerations

Ethical approval was granted by the UAE University Research Ethics Committee (no: ERSC_2024_5382), and the participating school's administration. Informed consent was obtained from all participants and their

Table 2. Paired-sample t-test results for pre- and post-test scores across Kilpatrick's strands

Kilpatrick's strand	Pre-test mean	Post-test mean	Mean gain	*t* (72)	*p*-value	Cohen's *d*
Conceptual understanding	1.0	2.5	1.5	7.58	< .001	0.88
Procedural fluency	4.5	8.5	4.0	9.46	< .001	1.10
Strategic competence	0.5	1.8	1.3	6.91	< .001	0.81
Adaptive reasoning	0.7	2.0	1.3	6.74	< .001	0.79

Table 3. Paired-sample t-test results for pre- and post-test scores across grade levels

Grade level	Pre-test mean (out of 25)	Post-test mean (out of 25)	Mean gain	*t* (72)	*p*-value	Cohen's *d*
Grade 5 (n = 28)	5.4	13.8	8.4	7.36	< .001	1.08
Grade 6 (n = 18)	7.6	16.8	9.2	8.12	< .001	1.25
Grade 7 (n = 15)	9.8	18.5	8.7	7.88	< .001	1.20
Grade 8 (n = 12)	11.2	19.7	8.5	6.93	< .001	1.10

guardians. Students were assured of voluntary participation and the right to withdraw at any time without academic consequence. All responses were anonymized through coded identifiers, and results were reported in aggregate form to preserve confidentiality. Data were stored securely and accessible only to the research team. The study adhered to the ethical standards for educational research involving minors, ensuring fairness, transparency, and participant welfare throughout all phases.

RESULTS

This section presents the findings of the six-week mathematics intervention, focusing on improvements in student performance across Kilpatrick's strands of mathematical proficiency and by grade level. Both descriptive and inferential statistics were employed to assess pre-test and post-test differences, with paired-sample *t*-tests determining statistical significance and Cohen's *d* used to estimate effect size.

Strand Level Performance

Table 2 displays the pre-test and post-test mean scores for each strand of mathematical proficiency. Results indicated statistically significant improvement across all strands. The largest gains were recorded in procedural fluency, where mean scores increased by four points (mean [M] = 4.5 to M = 8.5), t (72) = 9.46, p < .001, d = 1.10, representing a large effect. Conceptual understanding also showed substantial progress (M = 1.0 to M = 2.5), t (72) = 7.58, p < .001, d = 0.88. Moderate yet significant gains were observed in strategic competence (M = 0.5 to M = 1.8), t (72) = 6.91, p < .001, d = 0.81, and adaptive reasoning (M = 0.7 to M = 2.0), t (72) = 6.74, p < .001, d = 0.79.

The following section reports the findings of the six-week mathematics intervention, with a focus on improvements in student performance across Kilpatrick's strands of mathematical proficiency and by grade level. Descriptive statistics from pre-test and post-test scores were used to calculate mean performance and gain scores. **Table 3** summarizes performance by grade level. Each grade cohort demonstrated statistically significant improvement from pre-test to post-test. The most pronounced average gain occurred in grade 6 (M gain = 9.2, t [17] = 8.12, p < .001, d = 1.25), followed closely by grade 7 (M gain = 8.7, t (14) = 7.88, p < .001, d = 1.20), grade 8 (M gain = 8.5, t (11) = 6.93, p < .001, d = 1.10), and grade 5 (M gain = 8.4, t (27) = 7.36, p < .001, d = 1.08). These data are represented graphically in **Figure 1**.

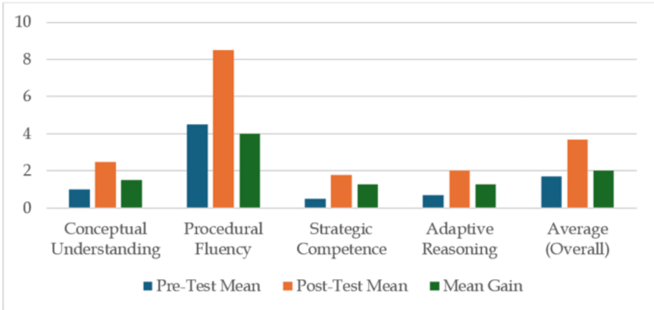

Figure 1. Strand-wise pre- and post-test results (Source: Authors' own elaboration)

Overall, students' mean mathematics scores improved by 8.7 points (p < .001, d = 1.05), indicating a large effect and substantial educational significance. Improvement patterns were consistent across grades and proficiency strands, validating the intervention's multidimensional impact on students' procedural, conceptual, and reasoning skills.

Qualitative Illustrations of Strand Level Change

To complement the quantitative gains reported above and to align the interpretation more clearly with the mathematical proficiency framework, we examined individual learners' pre- and post-test scripts. In **Figure 2** and **Figure 3** we present brief illustrative examples that show how changes in procedural fluency, conceptual understanding, and adaptive reasoning were manifested in students' work. The excerpts are representative of patterns observed across multiple learners rather than isolated cases.

In one item, students were required to solve a multi-digit subtraction problem with regrouping (e.g., subtracting a four-digit number from another four-digit

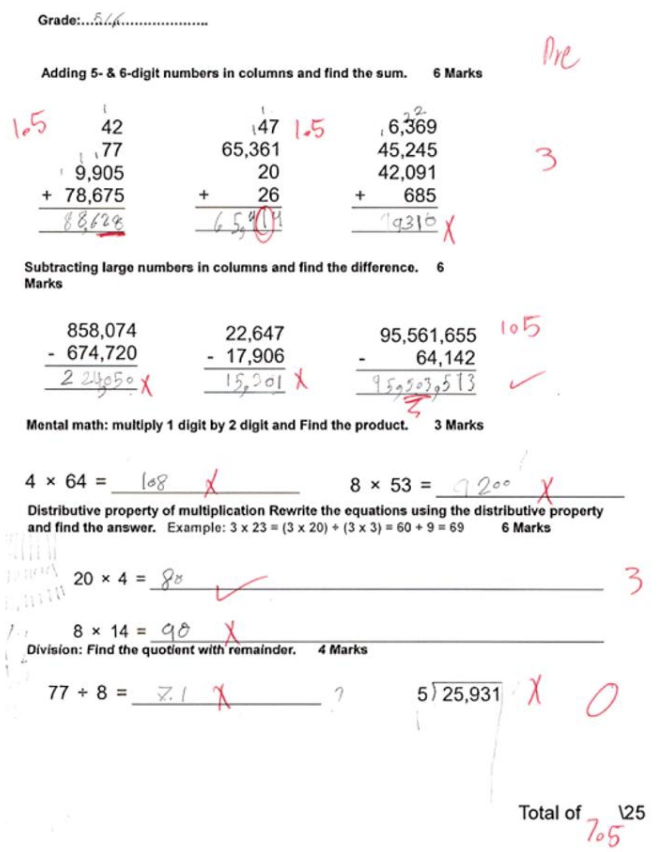

Figure 2. Pre-test results (Source: Field study)

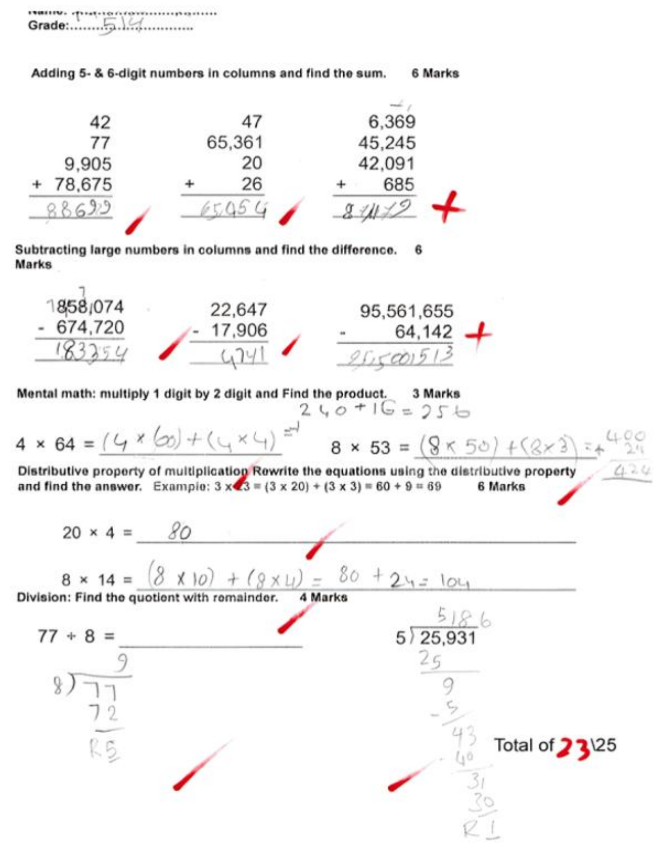

Figure 3. Post-test sample (Source: Field study)

number). On the pre-test, a grade 5 student recorded the numbers vertically but subtracted each column without regrouping, producing an incorrect answer and incomplete working. In the post-test, the same student correctly set up the subtraction, performed regrouping accurately across two place values, and showed each step clearly. This indicates improved procedural fluency, reflected in the more accurate and efficient execution of the standard algorithm. Another item asked students to use or explain the distributive property in a multiplication expression (e.g., rewriting $a \times b = a \times (b_1 + b_2)$ and explaining how the partial products relate to the original expression). In the pre-test, a grade 5 student simply rewrote the numbers or performed an unrelated calculation, with no indication of how the expression was being decomposed. In the post-test, the same student correctly decomposed one factor into tens and ones, wrote the corresponding partial products, and explained that "we break the number into parts, multiply each part, then add them back." This progression from an undeveloped response to a structured explanation illustrates a clearer conceptual understanding of the distributive property and its role in multiplication. In the pre-test, a grade 5 student obtained a numerical quotient but either ignored the remainder or rounded it without justification and provided no written explanation. In the post-test, the same student carried out the division correctly, stated the quotient and remainder, and explained in words how the remainder should be interpreted in the given situation. This shift from giving only an incomplete numerical answer to providing a justified interpretation indicates enhanced adaptive reasoning, as the student connects the numerical result to the real-world context. Taken together, these qualitative examples support the strand level quantitative findings. They suggest that students did not merely produce more correct answers; they increasingly demonstrated more accurate and efficient procedures, richer conceptual explanations, and more appropriate interpretations and justifications in their written work, consistent with improvements in procedural fluency, conceptual understanding, and adaptive reasoning.

DISCUSSION

The findings of this study reaffirm that targeted, strand-based interventions can produce substantial improvements in mathematical proficiency among low-achieving learners. The statistically significant gains across all four measured strands–procedural fluency, conceptual understanding, strategic competence, and adaptive reasoning–demonstrate the multidimensional nature of mathematical growth when instruction explicitly integrates Kilpatrick's framework. Students' overall mean improvement of 8.7 points ($p < .001$, $d = 1.05$) supports the argument that structured instructional approaches grounded in the five-strand model can yield rapid academic benefits even within short implementation periods.

These outcomes are best understood in light of the specific instructional activities undertaken during the intervention. Over the six weeks, lessons consistently combined conceptual warm-ups (for example, using decompositions and distributive-property explanations), structured practice of written algorithms for multi-digit operations, and strategy-based word problems that required students to justify and interpret their solutions, followed by whole-class reasoning discussions. Formative assessment routines such as warm-up questions, exit tickets, and quick response tasks were used to surface misconceptions and adjust pacing. The qualitative examples presented in the Results section make these strand level changes visible: students' multi-digit subtraction work shifted from incomplete or incorrectly regrouped procedures to fully worked, accurate algorithms (procedural fluency); their responses to distributive-property items evolved from undeveloped or unrelated calculations to coherent decompositions with verbal explanations of "breaking numbers into parts and multiplying each part" (conceptual understanding); and their division-with-remainder responses moved from bare numerical answers or ignored remainders to context-appropriate interpretations accompanied by written justifications (adaptive reasoning and strategic competence). Thus, the intervention did not simply increase the number of correct answers; it fostered richer forms of mathematical proficiency that are consistent with the underlying theoretical framework. In the context of TIMSS, where performance is assessed through the cognitive domains of knowing, applying, and reasoning, the present study provides evidence that progress across these areas can be achieved through balanced, strand-oriented teaching. By aligning the content of the intervention and the assessment with the TIMSS Number domain and cognitive processes, and by designing activities that explicitly linked procedures, concepts, and reasoning, the study demonstrates a plausible pathway through which classroom-level instruction can support "TIMSS readiness" for low-achieving learners. While the design does not permit direct claims about TIMSS score gains, the observed improvements in fluency, explanation, strategy use, and interpretation point to strengthened competencies that are central to success in large-scale assessments. Similar to findings from O'Dwyer et al. (2015) and Rittle-Johnson and Schneider (2014), the intervention's design facilitated connections between procedural and conceptual knowledge, enabling students to understand why algorithms work rather than simply how to execute them. The emphasis on discussing solution strategies and justifying procedures resonates with studies showing that opportunities for explanation and reflection can lead to concurrent gains in conceptual understanding and procedural fluency, particularly in whole-number operations. The results also align with global studies on mathematics interventions. For example, Betts et al. (2023) and Williams et al. (2022) found that structured middle-school interventions focusing on core arithmetic and reasoning tasks produced large effect sizes, particularly when instructional design combined modelling, feedback, and student reflection–components that were deliberately embedded in the present program through teacher modelling, guided practice, and whole-class reasoning discussions. However, unlike many large-scale interventions that span months or full academic years, this six-week study achieved comparable effect sizes within a significantly shorter duration, affirming the efficiency of strand-based instructional alignment. The sustained improvement across grades suggests that the approach can be flexibly adapted to different developmental stages, supporting similar conclusions by Adeniji and Baker (2022) regarding the scalability of conceptual–procedural integration strategies. The qualitative strand level evidence from students' written work strengthens this claim, as it shows that the gains are not limited to numerical scores but are reflected in more coherent representations, explanations, and solution plans. From a policy perspective, the findings have important implications for mathematics education in the UAE. National reforms under the Ministry of Education and the Emirates Schools Establishment have prioritized improving international assessment performance, particularly in TIMSS and PISA. Yet, as noted by Miao et al. (2021) and Ibrahim and Alhosani (2020), persistent gaps remain in conceptual reasoning and problem-solving–areas directly addressed by this intervention through its focus on distributive reasoning, interpretation of remainders, and multi-step word problems. The success of pre-service teachers in implementing the strand-based program highlights the potential of integrating this model into teacher preparation and ongoing professional development. Embedding Kilpatrick's five strands in curriculum planning, daily lesson design, and assessment practices can help ensure that classroom instruction aligns with the TIMSS cognitive progression from knowing to reasoning, thereby enhancing the nation's readiness for future benchmarking cycles. The observed improvement in students' mathematical engagement also reflects progress toward the UAE's vision 2031 and centennial 2071 educational goals, which emphasize critical thinking, innovation, and lifelong learning. By fostering productive disposition alongside cognitive proficiency, strand-based interventions can contribute to a more resilient and motivated learner population capable of approaching mathematics as a sense-making discipline. In this regard, classroom observations and student work samples indicating increased confidence, willingness to explain answers, and participation in discussions reinforce the notion that emotional and attitudinal factors are integral to sustainable achievement–an insight echoed by Boaler (2016) and Moussa and Saali

(2022). Despite its short duration, the intervention's large effect sizes suggest that even modest shifts in instructional design–such as routinely integrating conceptual warm-ups, explicit strategy sharing, and structured reasoning tasks–can catalyze significant learning improvements. At the same time, scaling such interventions requires systemic support, including time allocation within school timetables, access to manipulatives and digital tools, and continuous coaching for teachers to balance procedural and conceptual instruction effectively. Future implementations might also benefit from more fine-grained diagnostic pre-assessments that identify strand-specific weaknesses, allowing for even more targeted instruction and differentiated support.

In summary, the results demonstrate that strand-based pedagogy represents a viable and evidence-based pathway for strengthening TIMSS-related competencies in the UAE. When integrated into broader policy initiatives and teacher development frameworks, such interventions have the potential to close enduring performance gaps while nurturing a culture of mathematical thinking that extends beyond test performance toward genuine understanding, reasoning, and appreciation of mathematics as a coherent and purposeful domain. The combination of quantitative gains and qualitative evidence of strand level change offers a robust basis for future work seeking to scale and refine proficiency-oriented interventions in similar contexts.

CONCLUSION

This study provided empirical evidence that a short-term, strand-based mathematics intervention can meaningfully enhance the mathematical proficiency of low-achieving students in the UAE. The integration of Kilpatrick's five strands–conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition–proved effective in addressing both skill gaps and attitudinal barriers. Quantitative findings showed a statistically significant improvement in students' overall mathematics performance, with a mean gain of 8.7 points ($p < .001$, $d = 1.05$). These results confirm that structured, evidence-based instruction can foster measurable progress even within a condensed six-week timeframe. Beyond the immediate performance gains, the study's outcomes have broader implications for sustainable educational improvement. The findings also align with the UAE's national educational agenda, particularly the vision 2031 and centennial 2071 frameworks, which emphasize quality education, innovation, and global competitiveness. Developing students' reasoning, problem-solving, and adaptive thinking skills directly contributes to these national priorities. Implementing strand-based instruction on a larger scale can serve as a sustainable model for bridging the persistent gaps highlighted in international assessments such as TIMSS and PISA. By fostering both cognitive and affective growth–through the development of mathematical confidence, persistence, and curiosity–this approach supports the creation of resilient learners prepared to navigate the demands of a knowledge-based economy. Finally, the study highlights the potential for sustainable reform through small, targeted innovations that are scalable and replicable. Continued investment in teacher capacity building, assessment literacy, and curriculum alignment is essential to maintain the momentum generated by such interventions. If systematically adopted across schools, strand-based instruction can move mathematics learning in the UAE beyond short-term remediation toward a culture of deep understanding, reflective learning, and sustained academic excellence.

Limitations

The study's design imposes several limitations that warrant careful interpretation. First, the absence of a control group limits causal inference, as improvements cannot be attributed exclusively to the intervention. However, methodological safeguards–such as standardized testing conditions, equivalent test forms, and fidelity checks–helped mitigate internal validity threats. Second, the small sample size and focus on a single school reduce the generalizability of findings across the wider UAE population. Third, the short duration (six weeks) constrained the depth of conceptual and reasoning development achievable. Finally, the analysis relied on descriptive and inferential statistics but did not incorporate longitudinal or retention measures. Future research could address the limitations through larger, multi-school studies employing quasi-experimental or mixed-methods designs.

Recommendations

For educational practice and policy, the findings suggest that Kilpatrick's five strands should be systematically integrated into everyday mathematics instruction to ensure that conceptual understanding and reasoning receive equal emphasis alongside procedural practice. Teachers should be encouraged to extend such interventions over a longer period–ideally across a full term or academic year–to reinforce sustained gains in higher-order competencies such as reasoning, problem-solving, and adaptive thinking. Furthermore, incorporating strategies that nurture productive disposition, including goal-setting, self-reflection, and growth mindset activities, can strengthen students' confidence, motivation, and persistence in mathematics. Embedding strand-based pedagogy within teacher professional development and pre-service training programs is also essential for aligning classroom practices with assessment goals and ensuring coherence between instruction and student learning outcomes.

For future research, longitudinal studies are recommended to evaluate the durability of learning gains and their transfer to standardized assessments such as TIMSS. Expanding investigations across varied school settings, grade levels, and mixed-gender cohorts would enhance the generalizability of results and provide a broader understanding of the intervention's impact. Employing mixed-methods designs that combine quantitative achievement data with qualitative insights from classroom observations and student interviews could yield richer perspectives on both cognitive and affective outcomes. Moreover, developing and validating assessment tools capable of measuring all five strands–including productive disposition–through attitudinal surveys and performance-based tasks would strengthen the comprehensiveness and precision of future evaluations.

In conclusion, this study provides robust empirical evidence that short, targeted interventions grounded in Kilpatrick's strands can significantly enhance multiple dimensions of mathematical proficiency among low-achieving learners. By fostering procedural mastery, conceptual understanding, and reasoning ability in tandem, such approaches directly support improvement in TIMSS cognitive domains of *knowing*, *applying*, and *reasoning*. Embedding strand-based frameworks into mainstream mathematics education can help the UAE advance its national goal of cultivating mathematically proficient, confident, and globally competitive learners in alignment with vision 2031.

Author contributions: Both authors contributed equally to critical review, editing, and finalization of the manuscript and approved the final version.

Funding: This paper originated from a larger study based on a UAEU Start-up 2024 Research Grant. The funding is gratefully acknowledged.

Ethical statement: The authors stated that the study was approved by the UAE University Research Ethics Committee on 8 November 2024 with approval number ERSC_2024_5382. Written informed consents were obtained from the participants.

AI statement: The authors stated that generative AI was used to support language editing (grammar, clarity, and conciseness) of portions of the manuscript. No AI tool was used to generate or analyze data, produce results, or create/alter figures or tables. All content was reviewed, verified, and approved by the authors, who take full responsibility for the final manuscript.

Declaration of interest: No conflict of interest is declared by the authors.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the corresponding author.

REFERENCES

Adeniji, S. M., & Baker, P. (2022). Worked-examples instruction versus van Hiele teaching phases: A demonstration of students' procedural and conceptual understanding. *Journal on Mathematics Education, 13*(2), 337-356. https://doi.org/10.22342/jme.v13i2.pp337-356

Almarashdi, H. S., & Jarrah, A. M. (2023). Assessing tenth-grade students' mathematical literacy skills in solving PISA problems. *Social Sciences, 12*(1), Article 33. https://doi.org/10.3390/socsci12010033

Altarawneh, A. F., & Marei, S. T. (2021). Mathematical proficiency and preservice classroom teachers' instructional performance. *International Journal of Education and Practice, 9*(2), 354-364. https://doi.org/10.18488/journal.61.2021.92.354.364

Badri, M., Alnuaimi, A., Yang, G., & Rashedi, A. A. (2020). Examining the relationships of factors influencing student mathematics achievement. *International Journal of Innovation in Education, 6*(1), 12-30. https://doi.org/10.1504/ijiie.2020.106167

Balala, M. M. A., Areepattamannil, S., & Cairns, D. (2021). Investigating the associations of early numeracy activities and skills with mathematics dispositions, engagement, and achievement among fourth graders in the United Arab Emirates. *Large-Scale Assessments in Education, 9*, Article 13. https://doi.org/10.1186/s40536-021-00106-4

Betts, J. R., Zau, A. C., Bachofer, K., & Polichar, D. (2023). Changing the odds: Student achievement after introduction of a middle school math intervention. *Journal of Research on Educational Effectiveness, 17*(1), 65-88. https://doi.org/10.1080/19345747.2023.2170937

Burns, M. K., Walick, C. M., Simonson, G. R., Dominguez, L., Harelstad, L., Kincaid, A., & Nelson, G. (2015). Using a conceptual understanding and procedural fluency heuristic to target math interventions with students in early elementary. *Learning Disabilities Research and Practice, 30*(2), 52-60. https://doi.org/10.1111/ldrp.12056

Fraser, B. J., & Hasan, A. A. (2019). One-to-one tutoring and mathematics students' achievement in the United Arab Emirates. *Learning and Teaching in Higher Education Gulf Perspectives, 16*(1), 27-41. https://doi.org/10.18538/lthe.v16.n1.330

Ho, T. M. P. (2020). Measuring conceptual understanding, procedural fluency and integrating procedural and conceptual knowledge in mathematical problem solving. *International Journal of Scientific Research and Management, 8*(5), 1334-1342. https://doi.org/10.18535/ijsrm/v8i05.el02

Ibrahim, A., & Alhosani, N. (2020). Impact of language and curriculum on student international exam performances in the United Arab Emirates. *Cogent Education, 7*(1), Article 1808284. https://doi.org/10.1080/2331186x.2020.1808284

Jones, C. W. (2012). Building citizens for the Arab knowledge economy: Evidence from the United Arab Emirates. *Arab Knowledge Forum*. https://doi.org/10.18502/aqf.0106

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. National Academies Press. https://doi.org/10.17226/9822

Klang, N., Karlsson, N., Kilborn, W., Eriksson, P., & Karlberg, M. (2021). Mathematical problem-solving through cooperative learning: The importance of peer acceptance and friendships. *Frontiers in Education, 6*. https://doi.org/10.3389/feduc.2021.710296

Miao, X., Mishra, P. K., & Nadaf, A. (2021). Evidence and promises of AI predictions to understand student approaches to math learning in Abu Dhabi K12 public schools. *Gulf Education and Social Policy Review, 1*(2), 45-61. https://doi.org/10.18502/gespr.v1i2.8458

Mizyed, H. A., & Eccles, C. U. (2023). Understanding Emirati teachers' challenges in fostering problem-solving skills development in early years: A preliminary study. *Social Sciences & Humanities Open, 8*(1), Article 100561. https://doi.org/10.1016/j.ssaho.2023.100561

Moussa, N., & Saali, T. (2022). Factors affecting attitude toward learning mathematics: A case of higher education institutions in the Gulf Region. *SAGE Open, 12*(3). https://doi.org/10.1177/21582440221123023

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. TIMSS & PIRLS International Study Center, Boston College.

O'Connor, B. R. (2023). Methodologies to reveal young Australian *Indigenous students' mathematical proficiency. Mathematics Education Research Journal, 36*(2), 311-329. https://doi.org/10.1007/s13394-023-00447-z

O'Dwyer, L. M., Yang, W., & Shields, K. A. (2015). Teaching for conceptual understanding: A cross-national comparison of the relationship between teachers' instructional practices and student achievement in mathematics. *Large-Scale Assessments in Education, 3*, Article 1. https://doi.org/10.1186/s40536-014-0011-6

Pertiwi, A., & Wahidin, W. (2020). Are the mathematics textbooks for eighth-grade meeting the TIMSS 2019 mathematics framework? *Edumatika: Jurnal Riset Pendidikan Matematika, 3*(2), 129-142. https://doi.org/10.32939/ejrpm.v3i2.623

Rittle-Johnson, B., & Schneider, M. (2014). Developing conceptual and procedural knowledge of mathematics. In R. C. Kadosh, & A. Dowker (Eds.), *The Oxford handbook of numerical cognition* (pp. 1118-1134). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199642342.013.014

Serhan, D. (2014). Students' understanding of the definite integral concept. *International Journal of Research in Education and Science, 1*(1), 84-93. https://doi.org/10.21890/ijres.00515

Shana, Z., Naser, K., & Zeitoun, E. (2024). Impact of web-based learning platforms on primary school students' academic performance in the UAE: Exploring the digital frontier. Eurasia Journal of *Mathematics, Science and Technology Education, 20*(1), Article em2385. https://doi.org/10.29333/ejmste/14091

Singh, P., Hoon, T. S., Nasir, N. A. M., Han, C. T., Rasid, S. M., & Hoong, J. B. Z. (2021). Obstacles faced by students in making sense of fractions. *The European Journal of Social & Behavioural Sciences, 30*(1), 34-50. https://doi.org/10.15405/ejsbs.287

Subba, S. B., & Gotamey, H. K. (2022). The effects of the intervention program on low achiever students' learning achievement in classroom. *Journal of Education, Society and Behavioural Science, 35*(1), 58-70. https://doi.org/10.9734/jesbs/2022/v35i130399

UAE Ministry of Education. (2020). TIMSS 2019 school report–UAE overview. *United Arab Emirates*. https://web.khda.gov.ae/en/Resources/Publications/International-Assessments/Dubai-TIMSS-2019

Wardat, Y., Jarrah, A. M., Almassri, H., & Johnson, J. D. (2022). Assessing the impact of digital games-based learning on students' performance in learning fractions using (ABACUS) software application. *Eurasia Journal of Mathematics, Science and Technology Education, 18*(10), Article em2159. https://doi.org/10.29333/ejmste/12421

Williams, R. T., Citkowicz, M., Miller, D. I., Lindsay, J., & Walters, K. (2022). Heterogeneity in mathematics intervention effects: Evidence from a meta-analysis of 191 randomized experiments. *Journal of Research on Educational Effectiveness, 15*(3), 584-602. https://doi.org/10.1080/19345747.2021.2009072

Yeliz, Y. (2015). Sixth graders and non-routine problems: Which strategies are decisive for success? *Educational Research and Reviews, 10*(13), 1807-1817. https://doi.org/10.5897/err2015.2230

Yulian, V. N., & Wahyudin, W. (2018). Analysing categories of mathematical proficiency based on Kilpatrick's opinion in junior high school. *Journal of Physics: Conference Series, 1132*, Article 012052. https://doi.org/10.1088/1742-6596/1132/1/012052