An-Najah University Journal for Research – B

Humanities

مجلــــة جامعــــة النجــــاح للأبحــــاث – ب

العلوم الإنسانية

Enhancing Arithmetic Skills through an Augmented Reality Based Instructional Model: A Quasi-Experimental Study in Palestinian Primary Schools

Mohammed Khalil Odeh[1*], **Rahim Kouki**[2]

(Type: Full Article). Received: 05 March 2025, Accepted: 02 January 2026

Accepted Manuscript, First Online

Abstract: Objectives: This quasi-experimental study examined the effectiveness of an augmented reality (AR)–based instructional model in developing third graders' addition and subtraction skills in Palestinian primary schools. **Methods:** Ninety-seven students were assigned to a control group receiving conventional instruction ($n = 44$) and an experimental group receiving AR-based instruction ($n = 53$). Researcher-designed, curriculum-aligned pre/posttests assessed addition and subtraction proficiency. Group differences on the posttest were analyzed using independent-samples *t*-tests; a paired-samples *t*-test compared addition vs. subtraction within the experimental group. **Results:** The experimental group outperformed the control group on addition (Exp: $M = 2.76$, $SD = 0.89$; Ctrl: $M = 2.33$, $SD = 0.67$), $t(95) = 2.675$, $p = .009$, Hedges' g = 0.53; and on subtraction (Exp: $M = 2.58$, $SD = 0.80$; Ctrl: $M = 2.07$, $SD = 0.77$), $t(95) = 3.224$, $p = .002$, Hedges' g = 0.64. Within the experimental group, the difference between addition and subtraction subscores was not significant, $t(52) = 1.097$, $p = .278$, indicating comparable gains across operations. **Conclusions:** Sequencing AR lessons from virtual manipulatives to number-line representations and then symbolic work—with fading of supports—yields statistically significant and educationally meaningful improvements in core arithmetic.

Keywords: augmented reality, arithmetic, addition and subtraction, Grade 3, quasi-experimental, instructional design

تعزيز مهارات الحساب من خلال نموذج تعليمي قائم على الواقع المعزّز: دراسة شبه تجريبية في المدارس الابتدائية الفلسطينية

محمد خليل عودة[1]، رحيم الكوكي[2]،

تاريخ التسليم: 2025/3/5 تاريخ القبول: 2026/1/2

ملخص: الأهداف: تفحّصت هذه الدراسة شبه التجريبية فاعلية نموذج تعليمي قائم على الواقع المعزّز (AR) في تنمية مهارات الجمع والطرح لدى طلبة الصف الثالث الأساسي في المدارس الفلسطينيّة.

المنهجيّة: بلغ عدد المشاركين 97 طالبًا، وُزِّعوا إلى مجموعة ضابطة تلقّت تعليمًا تقليديًا (n =44) ، ومجموعة تجريبية تلقّت تعليمًا قائمًا على الواقع المعزّز (n = 53). استخدمت اختبارات قبليّة/بعديّة مُصمَّمة من الباحث ومواءَمة مع المنهاج لتقويم الكفاءة في الجمع والطرح. جرى تحليل الفروق بين المجموعتين على اختبار البعدي باستخدام اختبار (t) لعينتين مستقلّتين، كما قورنت درجتا الجمع والطرح داخل المجموعة التجريبية باستخدام اختبار (t) للعينات المترابطة.

النتائج: تفوّقت المجموعة التجريبية على الضابطة في أداء الجمع (التجريبية: M = 2.76, SD = 0.89؛ الضابطة: M = 2.33, SD = 0.67)، t(95) = 2.675, p = .009, Hedges' g = 0.53؛ وكذلك في أداء الطرح (التجريبية: M = 2.58, SD = 0.80؛ الضابطة: M = 2.07, SD = 0.77)، t(95) = 3.224, p = .002, Hedges' g = 0.64. وداخل المجموعة التجريبية لم يكن الفارق بين درجات الجمع والطرح دالًا إحصائيًا، t(52) = 1.097, p = .278، ما يشير إلى مكاسب متقاربة عبر العمليّتين. **الاستنتاجات:** يؤدّي تسلسل دروس الواقع المعزّز من المُجسّمات الافتراضيّة إلى تمثيلات خطّ الأعداد ثم العمل الرمزي—مع التخفيف التدريجي للدعم—إلى تحسينات دالّة إحصائيًا وذات قيمة تربويّة في الحساب الأساسي.

الكلمات المفتاحية: الواقع المعزّز؛ الحساب؛ الجمع والطرح؛ الصف الثالث الأساسي؛ تصميم شبه تجريبي؛ التصميم التعليمي.

---

[1] Ministry of Education and Higher Education – Nablus Directorate, Palestine .Orcid: https://orcid.org/0009-0004-9679-0541

* Corresponding author email: Mohammad.K.Odeh@ptuk.edu.ps

[2] Department of Laboratory of Art, Higher Institute of Childhood Education Executives (ISCE), Carthage, Tunisia.
Orcid: https://orcid.org/0000-0002-8664-731X

[1] وزارة التربية والتعليم العالي- مديرية نابلس، فلسطين

* الباحث المراسل: محمد خليل عودة

[2] مختبر بحوث الفنون، المعهد العالي لإطارات الطفولة، قرطاج، تونس
rahim.kouki@ipeiem.utm.tn

Introduction

Primary mathematics often struggles to secure durable understanding of addition and subtraction, where learners must coordinate place value, regrouping, and symbol–quantity mappings in real time. In the Palestinian context, system-level indicators underscore the challenge: in PISA 2022, 15-year-olds in the Palestinian Authority scored 366 in mathematics compared with the OECD average of 472, and only 20% reached at least Level 2 proficiency (Organisation for Economic & Development, 2023a). These figures point to persistent gaps in foundational numeracy and justify classroom-embedded interventions that target early operations with conceptually rich, engaging pedagogy.

Augmented reality (AR)—which overlays interactive digital objects onto the physical environment—has emerged as a practical way to make abstract number relations visible and manipulable for young learners. Two recent systematic reviews synthesize over a decade of studies in school mathematics and converge on design features associated with learning gains: deliberate sequencing of multiple, aligned representations (e.g., manipulatives → number lines → symbols) and immediate feedback that sustains attention and guides strategy formation (Ersen & Alp, 2024). Going beyond attitudes, a 2024 meta-analysis reports reliable AR benefits on ARCS motivational outcomes—attention, relevance, confidence, and satisfaction—design levers that matter when students learn multi-step procedures such as regrouping (Prasetya et al., 2024)

Evidence from primary-grade mathematics increasingly documents achievement and reasoning gains as well. For example, a 2025 classroom study on area measurement found that AR improved children's reasoning and—critically—maintained gains after supports were withdrawn, highlighting the value of staged scaffolds that fade over time, Hwang, & (Flavin et al., 2025). At the same time, a regional survey of VR/AR in Arab science and mathematics education shows that the evidence base remains sparse and unevenly distributed across countries, with few classroom-embedded studies at lower primary—underscoring the need for localized trials in Palestine (Abutayeh, Kraishan, & Kraishan, 2022 (Abutayeh et al., 2022); (Organisation for Economic & Development, 2023b).

Grounded in constructivism and cognitive-load principles, the present study evaluates a theory-informed AR model for Grade-3 arithmetic that maps theory to design: learners progress from virtual manipulatives to number-line representations and then to symbolic work, with just-in-time cues that fade as proficiency grows. In addition to group-level outcomes, the study differentiates by operation (addition vs. subtraction) to test where AR yields the greatest returns in foundational arithmetic.

Despite accelerating progress in AR-supported mathematics, three gaps remain salient for early arithmetic and for the Palestinian context. First—scope and setting: the Arab-region evidence base is sparse and unevenly distributed, with few classroom-embedded trials in lower primary grades, even as system-level indicators (e.g., PISA 2022) point to persistent shortfalls in foundational numeracy in the Palestinian Authority (Abutayeh et al., 2022)

Second—topic and method: recent systematic reviews converge on AR's benefits when tasks sequence multiple aligned representations with immediate feedback, but most studies skew toward secondary samples or non-arithmetic topics (e.g., geometry/measurement), and few disaggregate impacts by operation or adjust for baseline performance (Flavin et al., 2025) (İslim et al., 2024).

Third—design transparency: limited work operationalizes constructivism and cognitive-load management as binding design commitments—i.e., a principled progression from virtual manipulatives → number-line representations → symbolic work with just-in-time cues that fade—and then tests whether such sequencing yields comparable gains in addition vs. subtraction at Grade 3.

This study addresses these gaps by (a) implementing a classroom-embedded, quasi-experimental AR intervention in Palestinian primary schools; (b) explicitly mapping theory to design (constructivism and cognitive load to manipulatives/number-line/symbolic sequencing with planned fading); and (c) conducting operation-wise analyses (addition vs. subtraction) with baseline control and effect-size reporting. In doing so, it contributes localized, methodologically transparent evidence to an underrepresented setting while testing the conditions under which AR improves core arithmetic (İslim et al., 2024) (Organisation for Economic & Development, 2023a).

Research Problem

Despite encouraging international findings, empirical, early-grade AR studies situated in Palestinian classrooms are scarce. Conventional, textbook-led pedagogy often emphasizes rote procedures with limited visualization, leaving many learners without the conceptual supports needed for regrouping.

Meanwhile, recent AR syntheses indicate that sequenced multi-representation tasks with immediate feedback can enhance conceptual understanding and motivation ,(Ersen & Alp, 2024), yet classroom-embedded evidence at lower primary in Palestine is lacking. Accordingly, the study addresses the following problem:

To what extent does an augmented reality–based instructional model improve Grade-3 students' performance in addition and subtraction—relative to conventional instruction—in Palestinian primary schools?

Operationally, we test whether the AR sequence (manipulatives → number line → symbolic, with fading supports) yields statistically significant post-test advantages and whether these advantages are comparable across operations (addition vs. subtraction). This framing directly responds to current system indicators (Organisation for Economic & Development, 2023a) and aligns with recent design-focused AR reviews (Ersen & Alp, 2024; İslim et al., 2024)

Research Questions

To address the identified problem, the study seeks to answer the following research questions:

To address the stated problem, the study investigates the following questions at $\boldsymbol{\alpha} \leq$

.05 (two-tailed), with pretest performance statistically controlled where applicable:

1- Between-groups effectiveness Do students taught with the augmented reality (AR)–based instructional model score higher on the posttest than students taught with conventional instruction?
a. Does this advantage hold on the addition subscale?

b. Does this advantage hold on the subtraction subscale?

2-Within-AR comparison by operation**.** Among students in the experimental (AR) group, are posttest means for addition and subtraction significantly different?

Study Objectives

To comprehensively address the dimensions of the research problem, this study aims to achieve the following objectives:

1. **Design and implement** a theory-informed AR instructional sequence for Grade-3 arithmetic that maps constructivism and cognitive-load principles to practice (virtual manipulatives → number-line representations → symbolic work) with fading supports.
2. **Evaluate effectiveness** versus conventional instruction on posttest performance in **addition** and **subtraction**, reporting baseline-adjusted statistics and effect sizes.
3. **Examine operation-wise differences** within the AR condition to determine whether effects are comparable across addition and subtraction.
4. **Document implementation requirements** (teacher mediation, scaffold progression, classroom/infrastructure needs) to support scalable adoption.

Significance of the Study

The significance of this study lies in the expected outcomes, which may contribute to several areas:

Theoretical significance. The study operationalizes constructivism (enactive → iconic → symbolic) and cognitive-load management (signaling, segmentation, fading**)** as explicit design commitments, then links these to measurable outcomes by operation—strengthening the theory–design–evidence chain in early arithmetic.

Methodological/empirical significance. It provides classroom-embedded**,** quasi-experimental evidence from Grade-3 Palestinian settings on core operations, differentiates impacts by operation (addition vs. subtraction), and reports baseline-adjusted effects with transparent statistics—addressing gaps in local, early-grade AR research.

-Practical significance. The work yields a replicable teaching sequence (activities, scaffold tiers, fade-out schedule) and concrete guidance for teacher mediation and lesson orchestration, offering ready-to-use materials for schools and professional learning.

-Policy and implementation significance. Findings supply decision-relevant evidence for ministries and school networks on when and how to invest in AR-supported foundational numeracy (infrastructure, professional development, classroom management models), prioritizing implementations that demonstrably improve Grade-3 addition and subtraction.

Study Scope and Limitations

This study is limited by its scope in terms of geographic location, sample size, and its focus on a specific mathematical unit using AR technology.

- **Subject Matter:** The use of augmented reality in teaching a mathematics unit for third-grade students.

- **Geographical Scope:** The study was conducted at the Academy of the Holy Quran and Future Generation School, under the supervision of the Directorate of Education in Nablus, Palestine.
- **Participants:** A purposive sample of third-grade students.
- **Time Frame:** The study was conducted during the second semester of the (2023/2024)academic year.

Theoretical Framework and Previous Studies

The Concept of Augmented Reality in Education

Augmented reality (AR) blends virtual objects with the physical environment to create interactive, learner-centred experiences. In early mathematics, this affordance turns abstract ideas—quantity, place value, regrouping—into visible, touch-like actions as learners compose and decompose numbers, operate on base-ten blocks, and traverse number lines with dynamic cues.

From a constructivist stance, such activity supports the move from doing to seeing to symbolizing: children first act on quantities, then stabilize those actions through imagery, and finally encode them in written algorithms.

Complementing this, multimedia/cognitive-load research indicates that signaling (e.g., highlights on traded tens), segmentation (small, timed steps), and worked-example → problem transitions minimize extraneous processing while preserving the constructive work that builds schemas for regrouping and borrowing. In our study, these principles are enacted directly: lessons begin with manipulable units/tens and jump-wise number lines, are chunked into short, cued steps, and then fade hints and animations so that strategies, not scaffolds, are internalized (Bruner, 1966); (Mayer, 2020)

Constructivism and the Explicit Theory–Design Link

We treat constructivism not as background rhetoric but as a set of design constraints**.** First, tasks follow Bruner's enactive → iconic → symbolic progression: students drag and group virtual units and tens (enactive), see the same relations on an AR number line and place-value panels (iconic), and finally record the standard procedures (symbolic).

Second, guidance is just-in-time and fading**:** early sessions include stepwise hints and short animations; later sessions retain minimal visual signals; final sessions withdraw supports to require independent solution.

Third, we orchestrate social construction through brief teacher prompts ("How did you make a ten?" "What changed when you borrowed?") to surface and compare strategies (e.g., make-ten, compensation). Fourth, artifacts are co-located to reduce split attention and protect working memory, ensuring that cognitive effort targets numerical structure rather than interface management.

These decisions are the operational footprint of constructivism in the intervention; the expected mechanism of impact is a shift in strategy use (from counting-all to counting-on to decomposition/compensation) accompanied by stable or improved accuracy and reduced solution time once strategies consolidate(Bruner, 1966; Mayer, 2020)

Recent evidence strengthens these choices and helps tailor them to early arithmetic. A **2024** systematic review of 60 AR-in-mathematics studies reported consistent positive effects on achievement, motivation, and 3D thinking, while noting the preponderance of secondary-level samples—an opportunity to extend AR into lower primary (Ersen & Alp, 2024) A broader 2010–2024 review likewise concluded that the strongest gains occur when multiple, well-aligned representations are deliberately sequenced and feedback is immediate—exactly the features we foreground (Ersen & Alp, 2024)

A **2025** classroom study in *Educational Technology Research & Development* showed that an AR app for area measurement improved children's reasoning and that gains persisted after supports were withdrawn, validating our inclusion of a delayed-retention check and our insistence on fading rather than permanent scaffolds (Flavin et al., 2025)

Previous Studies

Recent scholarship has substantially expanded what is known about AR in school mathematics, and two 2024 reviews synthesize the state of the field while pointing to gaps directly relevant to early arithmetic. A decade-scope systematic review reported consistent, positive effects of AR on mathematics achievement, motivation, and 3D thinking, though with a notable tilt toward secondary samples—suggesting clear headroom to extend AR down to lower primary (Ersen & Alp, 2024). A broader 2010–2024 review converged on the design features that matter most—deliberate sequencing of multiple aligned representations and immediate feedback—precisely the ingredients our intervention operationalizes through virtual manipulatives, number-line supports, and just-in-time cues (İslim et al., 2024).

Reinforcing these design choices, a 2024 meta-analysis on motivational outcomes found reliable AR benefits on attention, relevance, confidence, and satisfaction when tasks adopt clear goals, tight step segmentation, and progressive challenge (Prasetya et al., 2024) e; these same elements underwrite our micro-tasks, correctness cues, and planned fading of hints.

Although many recent trials center on primary geometry and measurement, their mechanisms transfer cleanly to early operations. In a 2025 classroom study, an AR app for area measurement boosted elementary students' reasoning and—critically—sustained gains after supports were withdrawn, underscoring the value of fading and justifying our delayed-retention probe (Flavin et al., 2025)

Complementary 2024 work with primary pupils reported sizable improvements in spatial visualization with AR-based activities, indicating that animated, 3D cues can accelerate the concrete-to-symbolic shift that early arithmetic relies on (e.g., composing/decomposing tens). Building on this logic, research that targets arithmetic more directly shows that structured, stepwise guidance is especially powerful for vulnerable learners: an AR game for addition and subtraction with Grade-5 students with learning disabilities yielded significant post-test gains and high satisfaction (Sangsawang, 2023). In adjacent manipulative-based work, a virtual-manipulative package for subtraction with regrouping produced maintained accuracy when supports were explicitly faded, aligning with our own fade-out schedule and strategy-shift goals (Park et al., 2020). Together, these findings suggest that when AR sequences enactive–iconic–symbolic representations and prunes scaffolds over time, learners internalize regrouping rather than the prompts themselves.

At the regional level, a survey of VR/AR in Arab science and mathematics education documented sparse, small-scale implementations and limited attention to lower primary, highlighting the need for classroom-embedded, curriculum-aligned studies that examine core operations with rigorous designs (Abutayeh et al., 2022). The present study responds to that need by situating an AR intervention in Grade-3 addition and subtraction within the Palestinian curriculum, mapping constructivist tenets to concrete design moves (virtual base-ten blocks and number-line jumps; signaling; segmentation; fading), and evaluating not only accuracy and time but also strategy use (e.g., shifts from counting-all to counting-on to decomposition/compensation) and delayed retention. In comparison to prior work that skews toward older grades, non-arithmetic topics, special populations, or immediate post-tests only, our design-to-evidence chain is intended to test whether theoretically motivated scaffolds yield durable, strategy-level change in foundational arithmetic.

Methodology

Research Design

This study adopted a quasi-experimental design with a pre-test–post-test control group structure. This design is suitable for educational field settings where full randomization is often not possible. It allows for assessing the causal effect of an instructional intervention—in this case, augmented reality (AR) tools—on students' learning outcomes by comparing the performance of two groups under different instructional conditions.

The design enables the researcher to determine whether statistically significant differences exist in the development of addition and subtraction skills between the group exposed to the AR-based model and the group taught using traditional methods, while controlling for initial group differences through pre-testing.

Population and Sample

The study population comprised all third-grade students enrolled in Palestinian primary schools during the 2024–2025 academic year. A purposive, school-level sampling strategy was used to identify two private schools in Nablus—the Academy of the Holy Quran and the Future Generation School—on the basis of (a) comparable socio-academic profiles, (b) administrative approval to conduct classroom research, and (c) access to the devices required for the AR condition. Within these schools, intact Grade-3 classes were assigned at the class level to conditions to minimize cross-group contamination: one class received the augmented reality (AR)–based instruction (experimental, *n* = 44) and the other received conventional instruction matched for time-on-task (control, *n* = 53). Because assignment occurred at the class level under scheduling and resource constraints, the design was quasi-experimental.

All participants were male third-graders who met the following inclusion criteria**:** current enrollment in Grade 3, regular attendance, and parental/guardian consent with school authorization. Exclusion criteria were lack of consent or incomplete outcome data (students who transferred mid-study or missed the posttest were excluded from analysis). Both groups were taught by qualified mathematics teachers with comparable years of experience; teaching loads and content coverage were aligned to ensure equivalent exposure over the same calendar window. Baseline addition and subtraction performance was measured at pretest and used to inform subsequent analyses (e.g., as a covariate where applicable); descriptive pretest statistics are reported in **Table(1)** Ethical approval was obtained from the relevant school administration and the investigators' institutional review procedures.

Table 1. Sample Characteristics and Group Allocation

Group	n	Grade level	Gender	School source(s)	City / Country	Sampling strategy	Assignment unit	Instructional condition	Time-on-task parity	Academic year
Experimental (AR-based instruction)	53	Grade 3	Male	Academy of the Holy Quran; Future Generation School	Nablus, Palestine	Purposive (school-level)	Intact classes (class-level)	AR application for arithmetic (addition & subtraction)	Yes (matched calendar window and content scope)	2024–2025
Control (conventional instruction)	44	Grade 3						Textbook exercises and teacher-led board work	Yes (matched calendar window and content scope)	2024–2025

Research Tools and Data Collection

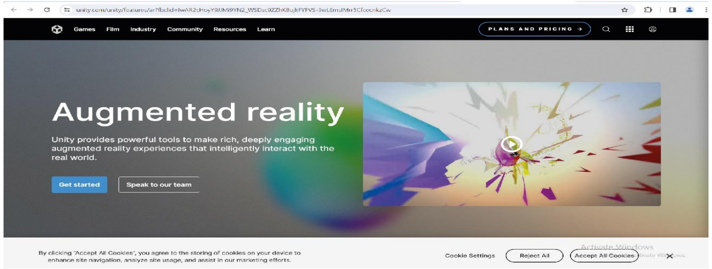
The study utilized:

1. An augmented reality-based instructional model.
2. A mathematical skills test (addition and subtraction).

.1 Augmented Reality Model Development

The study used the Unity AR Engine, a widely recognized development platform, to build the instructional application. The Unity-based app was made available at no cost on Google Play and the Apple App Store and enabled the integration of images, sound effects, and videos, which made the AR activities accessible and interactive. Given its affordability and ease of use, Unity proved effective for designing AR-based educational content.

For more details, visit: www.unity.com



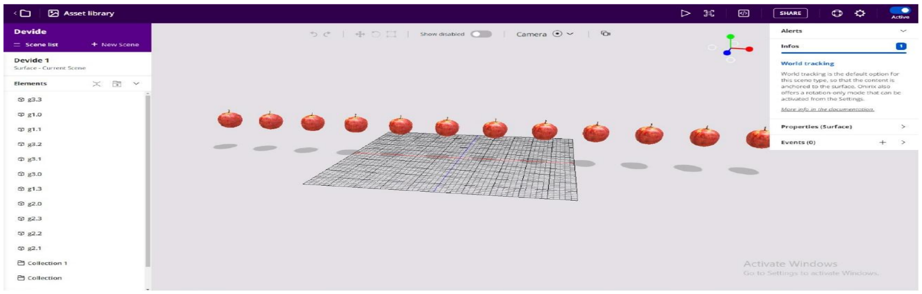
Augmented reality

Unity provides powerful tools to make rich, deeply engaging augmented reality experiences that intelligently interact with the real world.

By clicking "Accept All Cookies", you agree to the storing of cookies on your device to enhance site navigation, analyze site usage, and assist in our marketing efforts.

Steps for designing educational content using Unity AR Engine were implemented.

Step 1: 3D models of elements and examples were created.

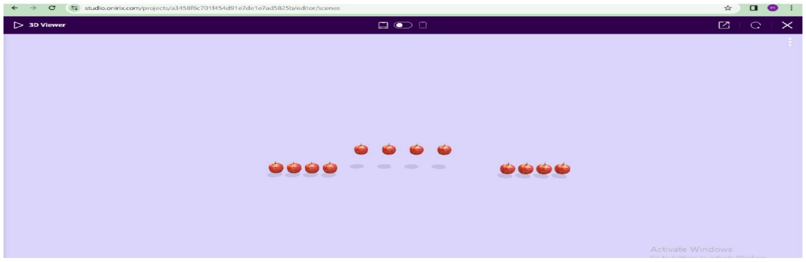


World tracking

World tracking is the default option for this scene type, so that the content is anchored to the surface. Onirix also offers a rotation-only mode that can be activated from the Settings.

More info in the documentation.

Step 2: Examples were programmed and their outputs were implemented in the AR application.



Step 3: Selected Curriculum Examples and Their Application to the Model

Examples were selected based on the attached patterns as follows:

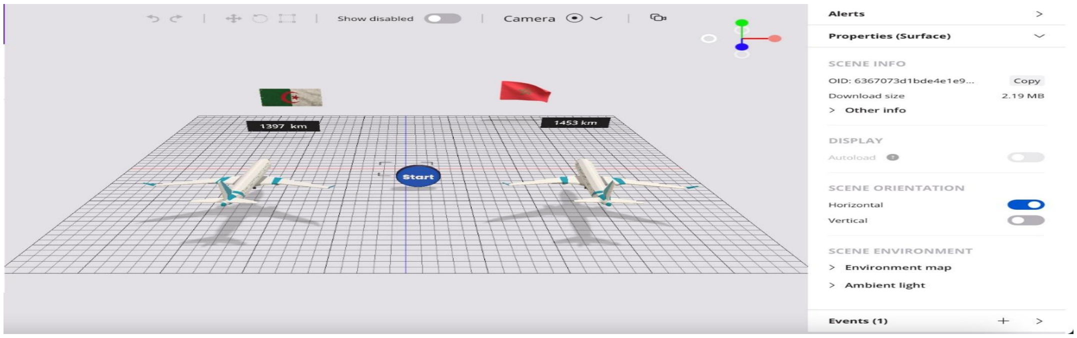
Model:

الدرسُ ٣ طرحُ عدديْن ضمنَ ٩٩٩٩٩

١ أَقلعتْ طائرتانِ تابعتانِ للخطوطِ الجويَّةِ الفلسطينيَّةِ من مطارِ ياسر عرفات الدوليّ في رفح، حيث كانتْ الأولى متجهةً إِلى المغربِ، بينما كانتْ الثانيةُ متجهةً إلى الجزائرِ، فقطعتِ الأولى مسافةَ ١٤٥٣ كيلو متراً، وقطعتْ الثانيةُ مسافة ١٣٩٧ كيلومتراً*.

أ ما مجموعُ المسافتيْنِ اللتين قطعتهما الطائرتان؟

The model was grounded in the first pattern (State Transformation), which relied on subtraction as the basic operation. The example was integrated into the augmented reality model using the following design:



OID: 6367073d1bde4e1e9...

Download size 2.19 MB

Horizontal ●

Vertical ○

Each Model’s Basis in Visual Communication and Didactic Approaches

Models were designed on visual-communication principles to deepen understanding of operations, classifications, and didactic situations. They were aligned with examples from the Palestinian Grade-3 mathematics textbook, and their interactivity employed a modern information-access strategy consistent with contemporary digital learning environments.

Validity of the Augmented Reality-Based Model

Content validity of the augmented reality (AR) instructional model was established through expert review. The model was evaluated by specialists in mathematics education, instructional design, and early childhood education, in addition to Grade-3 teachers and educational supervisors.

Experts assessed alignment with curriculum objectives, appropriateness for students' cognitive level, and clarity of mathematical content. Based on their feedback, necessary modifications were made.

To ensure consistency among expert evaluations, inter-rater agreement was calculated using Cooper's agreement formula. The results, presented in Table 2, indicate a high level of agreement across all evaluated components of the AR model.

Reliability of the Arithmetic Skills Test

The reliability of the arithmetic skills test, which measured students' addition and subtraction skills, was examined using Cronbach's alpha coefficient. The results indicated a high level of internal consistency for the overall test (α = 0.88), as well as for the addition subscale (α = 0.85) and the subtraction subscale (α = 0.84). These values suggest that the test provided reliable and consistent measurement of students' arithmetic performance.

Table 2: Inter-Rater Agreement on the Components of the Augmented Reality-Based Model

Component	**Number of Evaluators**	**Number of Items**	**Agreement Percentage**
Instructional Objectives	5	20	93%
Educational Content	8	20	95%
Learning and Assessment Activities	6	20	94%
Technical Design	4	20	85%
Overall Agreement	—	—	**91.75%**

Note. Agreement was computed using Cooper's Agreement Formula based on expert rating sheets

Intervention Schedule (Duration and Frequency)

The intervention was delivered over six instructional weeks. Students in the experimental group participated in two AR-integrated lessons per week, each lasting approximately 35 minutes, totaling 12 lessons and about 420 minutes of AR-supported instruction. The control group received instruction over the same weeks and lesson counts using conventional methods

Table 3: AR Intervention Dosage

Weeks	Lessons/week	Lesson length (min)	Total lessons	Total AR time (min)
6	2	35	12	420

Pre- and Post-Test for Arithmetic Skills

To evaluate the model’s effectiveness, a pre-test and post-test were developed to measure students’ proficiency in addition and subtraction. The tests were administered with a 90-day interval between administrations. The test development followed five steps: (1) a test blueprint was created specifying the number of instructional sessions per topic; (2) key mathematical topics were defined; (3) topic weights were calculated as (sessions for topic ÷ total sessions) × 100; (4) cognitive objective weights were calculated as (objectives at level ÷ total objectives) × 100; and (5) a table of specifications was constructed aligning content topics with cognitive targets to balance item types and difficulty.

Table 4: Content Analysis Framework for Arithmetic Concepts

Facts & Instructions	Values & Attitudes	Learning Outcomes	Skills	Concepts & Terminology
State transformation from part to whole	Collaboration in solving problems with peers	Student accurately sums numerical values	Summing similar place values; solving word problems	**Concept of Addition**
State transformation (Rearrangement)	Peer collaboration in exercises	Understanding and applying the **commutative property** of addition	Recognizing that **addition is commutative**	**Commutative Property of Addition**
Case Composition	Independent problem-solving	Accurately subtracting place values and applying borrowing when needed	Subtracting similar place values, borrowing, and identifying missing digits in subtraction	**Concept of Subtraction**

Pilot Study and Test Validation

The test was reviewed by curriculum specialists and mathematics educators for accuracy and grade-level appropriateness. Based on feedback, wording, structure, and sequencing were revised, and a final version was approved. A pilot test with 50 students was conducted to examine item difficulty, discrimination indices, and clarity; results informed further refinements. Internal consistency was assessed using Cronbach's alpha on the study sample.

Table 5. Internal Consistency (Cronbach's alpha) — provisional values

Scale / Subscale	Number of Items	Cronbach's α
Total test	20	0.88
Addition subscale	10	0.85
Subtraction subscale	10	0.84

Time Allocation for the Test

Test duration was determined using the midpoint rule: Duration = (time of first finisher + time of last finisher) / 2. Based on pilot timings (55 and 65 minutes), the duration was set to 60 minutes for both the experimental and control groups in the final administration.

Study Results

This section presents the empirical results using descriptive and inferential statistics. Results are organized by the research questions: RQ1a/b compares posttest means between the AR and control groups for addition and subtraction (independent-samples *t*), and RQ2 compares addition versus subtraction within the experimental group (paired-samples *t*).

Analytical note. RQ1 (between groups) was tested with an independent-samples t ($df = n_1 + n_2 - 2 = 95$ for 53 vs. 44), whereas RQ2 (within the experimental group only) was tested with a paired-samples t ($df = n - 1 = 52$ for $n = 53$). Accordingly, $p < .05$ denotes statistical significance; $p \geq .05$ denotes no statistically significant difference.

Results for the first question a–b (Between-Groups Posttests)

Independent-samples t-tests compared posttest means for the experimental (AR) and control groups on the addition and subtraction subscales (two-tailed, α = .05).

Table 6. Independent t-test results for posttest scores in addition and subtraction

Operation	Group	N	Mean (M)	SD	Df	t	P	Hedges' g [95% CI]
Addition	Control	44	2.33	0.67	95	-2.675	.009	0.53 [0.13, 0.94]
Addition	Experimental	53	2.76	0.89	—	—	—	—
Subtraction	Control	44	2.07	0.77	95	-3.224	.002	0.64 [0.24, 1.05]
Subtraction	Experimental	53	2.58	0.80	—	—	—	—

Note. df = 95 for between-group tests (n = 53 vs. n = 44). p-values are reported directly without asterisk notation

Results for the second question (Within-AR Operation-wise Comparison)

A paired-samples t-test compared experimental-group posttest means on addition and subtraction (two-tailed, $\alpha = .05$).

Table 7. Paired t-test comparing addition and subtraction posttests in the experimental group

Operation	n	Mean (M)	SD	df	T	P
Addition	53	2.76	0.89	52	1.097	.278
Subtraction	53	2.58	0.80	—	—	—

.*Note* Paired-samples *t* within the experimental group (n = 53); **df = 52 (= *n* − 1)**. The comparison was **not** statistically significant (*p* = .278).

Supplemental Effect-Size Diagnostics

Table 8. Supplemental effect-size metrics derived from reported tests

Comparison	R	ηp^2	Common-Language ES (CLES)
Addition: AR vs. Control	0.265	0.070	64.6%
Subtraction: AR vs. Control	0.314	0.099	67.5%
Within-AR: Addition vs. Subtraction	—	—	d_z ≈ 0.15

Note. r and ηp² derived from t and df; CLES expresses the probability that a randomly selected AR student outperforms a control student on the same subtest.

1 RQ1a–b: AR vs. Conventional Instruction on Posttests

1a Addition (AR vs. Control)

The experimental group outscored the control group on the addition subtest ($M_{(\text{exp})} = 2.76$, $SD = 0.89$; $M_{(\text{ctrl})} = 2.33$, $SD = 0.67$), $t(95) = 2.675$, $p = .009$, Hedges' $g = 0.53$ [0.13, 0.94]. As a rule of thumb in classroom research, $g \approx 0.50$ indicates a moderate, instructionally visible gain; expressed as a common-language effect, a randomly chosen AR student would outperform a randomly chosen control student about 65% of the time. In other words, the advantage is not only statistically reliable but also educationally meaningful under ordinary classroom conditions.

The intervention implemented Bruner's progression—enactive → iconic → symbolic—rather than merely citing it. Students first acted on virtual base-ten units/tens (enactive), then coordinated magnitudes on an AR number line (iconic), and only then encoded procedures symbolically (Bruner, 1966). This order matters: it lets learners construct place-value structure and shift strategies (from counting-on to composition/decomposition) before procedures harden into rote rules. From a cognitive-load stance, three design choices are pivotal: signaling (cues that keep attention on the relevant digits/places), segmentation (short, timed steps), and planned fading (supports that gradually recede). Together, these reduce extraneous load and split attention, preserving germane load for schema building (Mayer, 2020). The observed advantage in addition is exactly what such alignment predicts.

The result converges with primary-level interventions reporting higher arithmetic achievement and engagement with AR compared to conventional lessons and with reviews showing that sequenced, multi-representation tasks plus immediate feedback are associated with stronger mathematics outcomes (Ibáñez & Delgado-Kloos, 2018) (Ersen & Alp, 2024); (İslim et al., 2024). Meta-analytic evidence that AR boosts ARCS motivation (attention, relevance, confidence, satisfaction) offers a plausible motivational engine for persisting through multi-step practice (Prasetya et al., 2024). Importantly, classroom work indicating that gains persist when scaffolds are withdrawn (if fading is planned) matches our design logic (Flavin et al., 2025). The present finding adds context-specific evidence from Grade-3 classrooms in Palestine—an under-represented setting in early-grade AR research—showing that these mechanisms travel beyond the contexts most often studied.

Could novelty or teacher expectancy explain the gains? Time-on-task and content windows were matched across groups, and teachers had comparable experience—both steps that attenuate such threats. The similar SDs (0.89 vs. 0.67) argue against unstable variance. While richer designs (e.g., baseline covariates) would tighten causal claims, the magnitude and coherence of the pattern with theory and prior evidence make a purely artifactual account unlikely.

1b Subtraction (AR vs. Control)

The AR advantage was larger in subtraction ($M_{(\text{exp})} = 2.58$, $SD = 0.80$; $M_{(\text{ctrl})} = 2.07$, $SD = 0.77$), $t\,(95) = 3.224$, $p = .002$, $g = 0.64$ [0.24, 1.05]; $r = .314$, $\eta_p^2 = .099$. As a common-language effect, a randomly chosen AR learner would outperform a control peer about 67–68% of the time. In educational terms, this is a moderate-to-strong classroom effect.

Subtraction with regrouping has higher element-interactivity: the learner must track the minuend–subtrahend relationship, perform borrowing, and maintain place alignment within limited working memory. The AR sequence externalized the hidden steps via animated regrouping (e.g., trading a ten for ten ones) and co-located cues (arrows/highlights at the exact digit and place), cutting the split-attention penalty that burdens paper-only explanations. Once procedures stabilized, fading curtailed over-dependence on supports—so learners internalized the strategy rather than the prompt. This mechanism maps cleanly to load theory and predicts precisely where AR should help most: multi-step, cognitively demanding operations.

The pattern dovetails with task-specific work showing that explicit, stepwise, and faded supports improve accuracy and retention on regrouping (Park, Bouck, & Smith, 2020). Reviews similarly report larger effects when aligned representations are sequenced and feedback is immediate—conditions our design met (İslim et al., 2024) (Ersen & Alp, 2024) As with addition, primary-level interventions document achievement and engagement gains under AR (Ibáñez & Delgado-Kloos, 2018). The present result extends those findings by demonstrating that the regrouping bottleneck is especially responsive to AR designs that make intermediate states visible and then fade them.

Group dispersions were similar (0.80 vs. 0.77), supporting homogeneity. Because subtraction is multi-step, fidelity matters; the planned sequencing and brief fidelity checks reduce, though cannot eliminate, the risk that implementation artifacts inflated effects. A delayed-retention probe would test durability—a direction supported by prior AR-with-fading studies (Flavin et al., 2025).

RQ2: Within-AR Operation-Wise Pattern (Addition vs. Subtraction)

Within the experimental group, addition vs. subtraction posttests were not significantly different, t (52) = 1.097, p = .278, with a small paired effect ($d^z \approx 0.15$). After the same AR dose and sequencing, students reached comparable posttest levels across the two operations.

Subtraction is typically harder in Grade-3. Achieving parity suggests the design neutralized complexity by (i) animating regrouping so transient mental steps became visible, (ii) co-locating digits, base-ten blocks, and number-line positions to reduce split attention, and (iii) segmenting tasks into short micro-steps with just-in-time cues that faded as proficiency grew. This profile matches cognitive-load guidance (Mayer, 2020)and mirrors reports that faded, explicit supports help learners close the gap on regrouping (Park et al., 2020). It is also consistent with primary-level findings that well-designed AR can equalize performance on demanding topics by externalizing intermediate states (Ibáñez & Delgado-Kloos, 2018)

With n = 53, the study was underpowered to detect small within-condition differences; a null here is compatible with the larger between-group subtraction effect (g = 0.64 vs. 0.53). Future multi-site work with strategy-use coding (e.g., counting-on → decomposition/compensation) and delayed-retention checks can test whether parity persists once scaffolds fully disappear.

Results support a strong reading of Bruner's claim: when action on quantites (enactive) is mapped to visuals (iconic) before symbols, learners not only score higher but also narrow the difficulty gap between operations with different cognitive demands (Bruner, 1966). In short, it is the sequence—not the medium alone—that drives understanding.

The pattern underscores that signaling, segmentation, and planned fading are design constraints, not decorative features. By reducing extraneous load (e.g., split attention, transient information) and supporting germane load (schema formation for regrouping), AR can yield moderate, durable effects ($g \approx 0.5$–0.6) under real-classroom conditions (Mayer, 2020). Convergence with recent reviews (Ersen & Alp, 2024; İslim et al., 2024) and motivation meta-analyses (Prasetya et al., 2024) suggests that alignment—multiple representations + immediate feedback + fading—explains as much (or more) of the gain as AR's novelty.

Beyond agreement with international findings (Ibáñez & Delgado-Kloos, 2018) the study addresses a documented gap in Arab-region, early-grade, classroom-embedded AR trials (Abutayeh et al., 2022) by providing localized Grade-3 evidence from Palestinian schools.

Across RQ1a (addition), RQ1b (subtraction), and RQ2 (within-AR comparison), a theory-informed AR sequence—manipulatives → number line → symbolic, with signaling/segmentation and planned fading—produced statistically significant, educationally meaningful advantages over conventional instruction while neutralizing the usual complexity penalty of subtraction. The pattern triangulates with constructivist and cognitive-load predictions and converges with recent syntheses and classroom studies. In practical terms, when AR is designed as a learning sequence rather than a stand-alone novelty, it can shift both accuracy and strategy sophistication in foundational arithmetic in ways that are theoretically legible and classroom-replicable.

Study Recommendations

For classroom practice (Grade 3 arithmetic).

- **Adopt an AR lesson sequence grounded in theory:** manipulatives → number line → symbolic recording, with signaling, segmentation (short, timed steps), and planned fading of supports.
- **Dosage guideline:** implement 2 AR-integrated lessons/week × 35 minutes for 6 weeks (≈ 12 lessons / 420 minutes) aligned to addition and subtraction with regrouping.
- **Fidelity tools:** use a one-page checklist (enactive/iconic/symbolic present; cues/fading applied; immediate feedback delivered) and a quick exit-ticket (2 items) to verify the targeted strategy (e.g., decomposition/compensation) was used—not guessed.
- **Formative assessment:** embed item-level feedback and two "no-hint" items per lesson to monitor internalization as supports fade.
- **Equity and access:** provide offline-capable AR content on shared devices; rotate stations so classes with limited hardware still complete the full sequence.

For teacher professional development.

- **Micro-PD cycle (3×60 minutes):** (1) modeling the AR sequence and error patterns in regrouping, (2) guided rehearsal with peer feedback using the fidelity checklist, (3) data use—reading item diagnostics and planning next-lesson adjustments.
- **Design principles:** emphasize how **constructivism** (enactive→iconic→symbolic) and **cognitive load** (reduce split attention; manage extraneous load; protect germane load) translate into concrete lesson moves (what to cue, when to fade).
- **Coaching & support:** pair teachers to co-plan one AR lesson per week and conduct brief 5-minute "learning walks" focused on signaling and fading.

For school leaders and policymakers.

- **Procurement with learning criteria:** favor AR tools that (i) allow multi-representation sequencing, (ii) provide immediate, item-level feedback, and (iii) support usage analytics (time-on-task, hint use, error types).
- **Monitoring targets:** track cohort gains of ~0.5 SD on curriculum-aligned mini-tests by week 6 and include a 2–3-week retention check (no hints) to verify durability.
- **Data protection:** adopt a minimal-data policy for young learners and require offline mode by default; ensure parental information sheets clarify purpose and data handling.
- **Scaling:** begin with Grade 3 arithmetic, then extend to adjacent strands (place value, multi-digit addition/subtraction) after a term of stable implementation.

For curriculum and assessment teams.

- **Curricular alignment:** map textbook examples to AR tasks that make intermediate states visible (e.g., animated regrouping) and then require no-scaffold practice.
- **Assessment design:** incorporate strategy-use items (e.g., "show your regrouping step") and a short

delayed-retention probe to check transfer beyond immediate practice.

Conclusion

This study found that an AR-based instructional sequence produced statistically significant, educationally meaningful advantages over conventional instruction in Grade-3 arithmetic. On posttests, the AR group outperformed the control group in addition (Hedges' $g \approx \mathbf{0.53}$) and **s**ubtraction ($g \approx \mathbf{0.64}$), with common-language effects of ≈ 65–68% favoring AR. Within the experimental group, addition vs. subtraction differences were not significant**,** indicating that the design neutralized the typical complexity penalty of subtraction through animated regrouping, co-located cues, and planned fading. These results are consistent with constructivist and cognitive-load accounts and extend the literature by offering classroom-embedded evidence from Palestinian primary schools. When AR is implemented as a learning sequence—not a novelty—it can meaningfully improve **accuracy** and strategy sophistication in foundational arithmetic.

Limitations and International Relevance

Limitations:

- **Sampling and context:** the sample was male, Grade-3, from two private schools in Nablus**;** findings may not generalize to mixed-gender or public-school settings.
- **Design scope:** the primary confirmatory analyses were posttest comparisons (between groups) and paired posttest comparisons (within AR). Without baseline covariates by operation, residual pre-existing differences cannot be fully ruled out.
- **Duration and follow-up:** the intervention ran for one half-term; there was no delayed-retention posttest in the main analyses, so long-term durability requires further verification.
- **Implementation variance:** although a fidelity checklist was used, micro-variations in pacing, cue timing, or fading may have introduced classroom-level noise.

International relevance:

Despite these constraints, the pattern—moderate gains in addition and larger gains in multi-step subtraction—is portable to systems facing similar early-numeracy challenges**.** The core design rules (multi-representation sequencing, immediate feedback, signaling/segmentation/fading) are platform-agnostic and feasible with low device ratios via station-rotation and offline content. For ministries and networks seeking scalable improvements in foundational math, the findings support targeted AR integration beginning in Grade 3, coupled with micro-PD and simple fidelity tools**.** Multi-site replications with mixed-gender and public-school samples, plus retention checks and strategy-use coding**,** would strengthen generalizability and inform policy at scale.

Ethics approval & consent: The study was conducted in accordance with institutional ethical standards. Administrative permissions were obtained from participating school.

Data availability: Data are available from the corresponding author upon reasonable request.

Conflict of interest: The authors declare no conflict of interest.

Author contribution: All authors contributed equally to this work in all sections

Funding: This research received no external funding.

Acknowledgment: The author extends sincere gratitude to Academy of the Holy Quran and Future Generation School for their generous cooperation in facilitating the implementation of this study. The author also expresses appreciation to the teaching staff, particularly the coordinating teachers, as well as the administrative teams in both schools, for their valuable support and assistance, which significantly contributed to the successful execution of the research. The author further acknowledges all individuals who contributed to the completion of this study and engaged with its instruments. In addition, the author expresses deep appreciation to the editorial board of *An-Najah National University Journal for Research* for their professional handling, continuous follow-up, and distinguished peer-review process, which contributed to presenting the study in its best scholarly form.

Open Access

References

115–139. https://doi.org/10.17275/per.24.52.11.4

Mayer, R. E. (2020). *Multimedia learning* (3 ed.). Cambridge University Press.

Organisation for Economic, C.-o., & Development. (2023a). *PISA 2022 results (Vol. 1): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

Organisation for Economic, C.-o., & Development. (2023b). *PISA 2022 results: Country note – Palestinian Authority*.

Park, J., Bouck, E. C., & Smith, R. (2020). Using virtual manipulatives to support elementary students with and without disabilities in subtraction with regrouping. *Journal of Special Education Technology*, *35*(4), 215–228. https://doi.org/10.1177/0162643419879252

Prasetya, P. A., Hartanti, D., & Wibowo, F. C. (2024). The impact of augmented reality learning experiences based on the motivational design model (ARCS): A meta-analysis. *Social Sciences & Humanities Open*, *10*, 100520. https://doi.org/10.1016/j.ssaho.2024.100520

Sangsawang, T. (2023). The use of augmented reality games in teaching addition and subtraction of numbers up to a hundred thousand for learning disability students in Grade 5. *Turkish Online Journal of Educational Technology*, *22*(1), 123–134.

Abutayeh, K. A., Kraishan, O. M., & Kraishan, G. M. (2022). The use of virtual and augmented reality in science and math education in Arab countries: A survey of research (2012–2021). *Frontiers in Education*, *7*, 979291. https://doi.org/10.3389/feduc.2022.979291

Bruner, J. S. (1966). *Toward a theory of instruction*. Harvard University Press.

Ersen, Z. B., & Alp, Y. (2024). A systematic review of augmented reality in mathematics education in the last decade. *Malaysian Online Journal of Educational Technology*, *12*(1), 15–31. https://doi.org/10.52380/mojet.2024.12.1.476

Flavin, M., Chung, J., Hwang, G. J., & Flavin, B. (2025). Augmented reality in primary mathematics: Conceptual reasoning, scaffolding, and retention in area measurement. *Educational Technology Research and Development*, *73*(4), 2663–2697. https://doi.org/10.1007/s11423-025-10502-0

Ibáñez, M.-B., & Delgado-Kloos, C. (2018). Augmented reality for STEM learning: A systematic review. *Computers & Education*, *123*, 109–123. https://doi.org/10.1016/j.compedu.2018.05.002

İslim, Ö. F., Namlı, Ş., Sevim Çırak, N., Özçakır, B., & Lavicza, Z. (2024). Augmented reality in mathematics education: A systematic review. *Participatory Educational Research*, *11*(4),