

A Learner-Generated Input Framework for AI and AR-Enhanced Primary Education: Design and Pilot Evaluation

Abstract— Primary-school geometry and foundational literacy present different learning problems, but both can benefit from an interaction cycle in which learner-generated input is computationally interpreted before feedback and three-dimensional (3D)/augmented-reality (AR) reinforcement is presented. This paper evaluates an Android educational suite that instantiates this common architecture through four components: Component 1 (C1), 2D-to-3D Geometry Validation; Component 2 (C2), 2D Net Correction and Virtual Slicing; Component 3 (C3), camera-based AI Air-Writing Trajectory Tracking; and Component 4 (C4), Sketch-Driven Multisensory Learning. The geometry components C1 and C2 were evaluated with learners aged 9-12. The literacy components C3 and C4 were evaluated with learners aged 6-8. A total of 50 different learners participated in the study: 25 learners aged 9-12 evaluated both C1 and C2, and 25 learners aged 6-8 evaluated both C3 and C4. C1 was tagged with the expected validity in 25/25 net trials; C2 completed successfully 20/25 Virtual Slicing sessions; C3 correctly recognized 19/25 evaluated air written letters; and C4 completed the full spelling, pronunciation-oriented and AR-reward sequence in 17/25 sessions. Average pre-test/post-test scores increased from 5/10 to 9/10 for C1, C2 and C3 and from 4/10 to 8/10 for C4. Results provide descriptive evidence that a common learner-input, analysis, feedback, and reinforcement architecture can support age-appropriate geometry and literacy activities while maintaining component-specific processing and evaluation.

Index Terms—air-writing recognition, augmented reality, geometric net validation, sketch recognition, virtual slicing.

I. INTRODUCTION

A. Background, Context, and Integration Rationale

Augmented reality can make abstract learning content directly inspectable in the learner's environment, but its educational value depends on how interaction is designed rather than on 3D visualization alone. Recent meta-analyses report positive average effects for AR-supported learning while also emphasizing variation by task, pedagogy, and implementation [1]. K-12 reviews similarly highlight the need to align usability and learning design with learner age and context [2]. These considerations are particularly relevant to primary geometry, where learners must coordinate flat representations with solid forms, and to early literacy, where learning involves ordered movement, visual-symbol association, spelling, and spoken-word practice.

Geometry and literacy are therefore not combined in this work because they use the same recognition algorithm. They are combined because all four components implement the same learner-centered interaction contract: learner-generated input is captured, interpreted by domain-specific logic, converted into an instructional state, and returned as targeted feedback followed by 3D/AR reinforcement. In geometry, the input is a net or slicing action; in literacy, it is a free-space letter trajectory or object sketch. This shared architecture provides a coherent research contribution while allowing

each component to retain its own algorithm, data representation, and evaluation measure.

B. Research Problem and Gap

Recent AR geometry studies show improvements in spatial visualization and geometric understanding [3]–[7], but they predominantly begin from predefined objects or instructional scenes. They do not generally verify whether a learner-created 2D net is mathematically foldable before 3D reinforcement, nor do they combine fault-specific correction with learner-controlled slicing of solids. Sketch-to-3D and sketch-recognition research demonstrates that freehand input can be computationally interpreted [8]–[10], but reconstruction or classification alone does not address geometric validity or educational correction.

Air-writing research demonstrates that fingertip trajectories can preserve movement information that static images discard [11]–[15]. In parallel, recent AR language-learning studies report value for alphabet and vocabulary activities [16]–[18]. However, these strands usually optimize a single recognition or visualization task. The unresolved problem addressed here is how learner-generated input can be computationally validated or recognized before reinforcement is presented across two age-appropriate learning pathways. The study therefore tests a common learner-input–analysis–feedback–reinforcement architecture rather than claiming novelty merely from combining unrelated features.

C. Aim, Objectives, and Research Questions

The aim is to design and evaluate an integrated mobile AI/AR educational suite that supports learner-generated geometry and literacy activities through targeted analysis, feedback, and 3D/AR reinforcement.

- 1) Validate learner-created 2D nets and transform valid constructions into interactive 3D solids.
- 2) Identify geometric faults, guide correction, and support internal/fractional reasoning through Virtual Slicing.
- 3) Analyze camera-tracked uppercase and lowercase English air-writing trajectories with attention to path, start position, stroke sequence, and direction.
- 4) Connect learner-generated object sketches with recognition, 3D visualization, spelling, pronunciation-oriented interaction, and AR reinforcement.

Six research questions guide the evaluation:

RQ1: How accurately and reliably does C1 distinguish valid and invalid learner-created nets and progress valid nets to the intended 3D state?

RQ2: How reliably does C2 execute fault-specific geometry support and the intended Virtual Slicing interactions across supported shapes, directions, and fractions?

RQ3: How accurately and responsively does the camera-based Air-Writing component recognize learner-generated letter trajectories?

RQ4: How reliably does the sketch-driven vocabulary component execute recognition, spelling, pronunciation-oriented, and AR-reward states?

RQ5: To what extent do learners' pre-test and post-test scores change after interaction with the age-appropriate components?

RQ6: How usable and understandable is the integrated mobile experience based on observed task completion, retry behavior, and completion time?

D. Research Contributions

The principal contribution is a reusable mobile interaction architecture in which learner-generated input is analyzed before reinforcement is shown. Four specialized mechanisms instantiate it: deterministic geometric-net validation; fault-specific correction and Virtual Slicing; camera-based air-writing trajectory recognition; and sketch-driven multisensory vocabulary learning. The integration is therefore architectural and pedagogical, while the algorithms and evaluation metrics remain component specific.

II. RELATED WORK

A. AR Geometry and Spatial Learning

Primary-school AR geometry studies report benefits for spatial visualization and geometric thinking [3]–[5], while intelligent AR tutoring and recent comparative work show that the effectiveness of 3D representations depends on task design and how learners manipulate them [6], [7]. These studies establish the value of interactive 3D geometry but mainly operate on predefined content. The present geometry pathway instead requires the learner's own net to pass a mathematical validity check before 3D conversion and uses correction and slicing as explicit learning states.

B. Learner-Generated Geometry, Correction, and Slicing

Sketch2Model demonstrates view-aware generation of 3D meshes from freehand sketches [8], and recent sketch-recognition work shows that CNN-based representations can classify hand-drawn forms [9], [10]. These approaches target reconstruction or recognition, whereas C1 treats a geometric net as a constrained mathematical object. C2 then uses the invalid state to explain a detected relation and, in its Virtual Slicing stage, allows vertical, horizontal, or angled cuts at $1/2$, $1/3$, $1/4$, or $3/4$. Recent AR work on cross-sections and geometric manipulation supports the pedagogical value of making internal structure visible [4], [7], but the closest systems do not combine learner-net validation, fault-specific correction, and slicing in one mobile workflow.

C. Camera-Based Air-Writing and Trajectory Recognition

Vision-based air-writing tracks a fingertip in free space and converts its ordered motion into a recognizable trajectory. Webcam-based work has demonstrated real-time fingertip tracking [11], MediaPipe Hands enables on-device hand-landmark estimation from a single RGB camera [12], and trajectory-based CNN/LSTM approaches show that normalization and sequential features are effective for character recognition [13], [14]. Recent educational air-writing systems also combine hand tracking with corrective feedback [15]. C3 applies this direction to English letter

learning for young children and explicitly treats the movement path—not only the final character image—as the instructional evidence.

D. Sketch Recognition and Multisensory Vocabulary Learning

The C4 learning goal is vocabulary rather than formal phonics. It connects a learner's object sketch with the corresponding English word, spelling, pronunciation-oriented practice, and AR reinforcement. Recent studies report AR-supported gains or motivation for alphabet and vocabulary learning [16], [17], and a 2024 review identifies interactivity and form–meaning mapping as recurring affordances in AR/VR vocabulary research [18]. C4 differs by using the learner's own sketch as the entry point instead of a predefined vocabulary item.

TABLE I. CLOSEST-WORK POSITIONING

Closest work	What it demonstrates	Differences in this study
Nadzeri et al. [3]	AR improves primary spatial visualization.	C1 validates learner-created nets before 3D output.
Uriarte-Portillo et al. [6]	AR + tutoring supports geometry.	C2 uses detected net faults and learner-controlled slicing.
Watanabe et al. [14]	Camera hand pose + deep model recognizes air-writing.	C3 uses camera-tracked trajectories for child letter-learning feedback.
Korosidou [16]	AR supports young learners' alphabet/vocabulary work.	C4 begins with learner-generated sketches and gates AR reward through tasks.

III. RESEARCH METHODOLOGY

A. Research Design and Target Users

The study follows Design Science Research (DSR): the integrated mobile suite is treated as an artifact that is iteratively designed and evaluated in context [19]. To strengthen evaluation, technical evidence is separated from learner outcomes, consistent with recent calls to combine design-science artifact evaluation with behavioral evidence [20]. The target population is divided into two age bands rather than treated as a single 6–12 group. C1 and C2 were evaluated with learners aged 9–12 because nets, solids, internal geometry, and fractional slicing require

comparatively stronger spatial reasoning. C3 and C4 were evaluated with learners aged 6–8 because letter formation, object–word association, basic spelling, and spoken-word practice are foundational literacy activities. The study was conducted following institutional ethics procedures. Parental/guardian informed consent and learner assent were obtained before participation, and all participant data were anonymized.

B. Implementation and Data Strategy

The application is implemented in Unity 2022.3 LTS and C# for Android. AR Foundation with the ARCore XR Plugin provides horizontal-plane detection, raycasting, placement, and physical-viewpoint inspection. C1 is deterministic and uses shape-specific rules over faces, edges, dimensions, connectivity, adjacency, and orientation. Supported geometry is Cube, Cuboid, Triangular Prism, Tetrahedron, and Cylinder. C2 consumes invalid geometry states for correction and supports the same five solids for Virtual Slicing.

C3 is a genuine camera-based air-writing interaction rather than touchscreen tracing. The mobile camera captures the learner's hand; hand-landmark tracking follows the index fingertip to form an ordered trajectory in free space. The documented recognition pipeline is CNN-based. Trajectory preprocessing normalizes the captured path before classification so that recognition is less sensitive to writing location and scale. The learner evaluation log records target and predicted letter, confidence, trajectory point count, stroke count, attempts, errors, drawing duration, and total session duration. The implemented scope supports uppercase and lowercase English letters; the final child-user dataset used in this paper contains 25 recorded trials and samples 17 uppercase target letters, so the results below are restricted to those evaluated trials rather than generalized to the full alphabet.

C4 uses Python, TensorFlow/Keras, OpenCV, FastAPI, NumPy, and Pillow during model development and Unity integration. The sketch classifier covers Apple, Cat, Fish, Sun, and Tree. Its development dataset contains 25,000 Quick, Draw!-derived images (5,000 per class), split into 17,500 training, 3,750 validation, and 3,750 held-out test images. OpenCV preprocessing normalizes the recognition input while Unity retains drawing bounds for visual correspondence. The final user evaluation additionally includes 25 sketches produced by children aged 6–8 in the actual application, addressing the gap between benchmark-style data and target-user drawings.

C. Evaluation Design

The evaluation involved 50 unique learners. The same 25 learners aged 9–12 evaluated C1 and C2 and the same 25 learners aged 6–8 evaluated C3 and C4. Independent analysis of technical and workflow results was done: C1 compares system validity decisions to stored expected labels; C2 analyzes success of slicing, direction/fraction state, tracing and completion time; C3 analyzes target vs. predicted character results and timing; and C4 analyzes spelling completion, pronunciation-oriented acceptance, AR unlocking, model confidence, attempts and session time. Educational change was assessed with 10-mark pre- and post-tests. Because only aggregate pre-/post-test means are retained for this paper, the learning results are descriptive, and

no inferential significance claim is made. Component-level technical and workflow evidence is summarized in Table II, while the age-specific pre-test and post-test learning outcomes are reported separately in Table III.

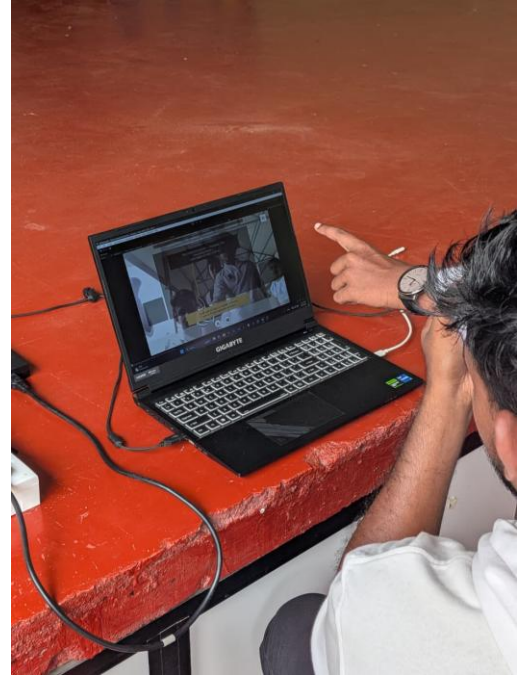


Fig.1. User Testing Evidence

IV. SYSTEM ARCHITECTURE AND COMPONENT DESIGN

A. Shared Architecture

Integration occurs at the interaction and presentation levels rather than by forcing the four components into one recognition model. Each pathway follows the same contract: capture learner input -> interpret it with the relevant component -> generate an instructional state -> return corrective, progression, or reinforcement feedback through the shared Unity interface. C1 and C2 are tightly coupled because an invalid C1 state routes to correction, whereas C3 and C4 use separate literacy inputs but share navigation, 3D assets, and AR presentation. Fig.2 and Fig.3 summarize this architecture.

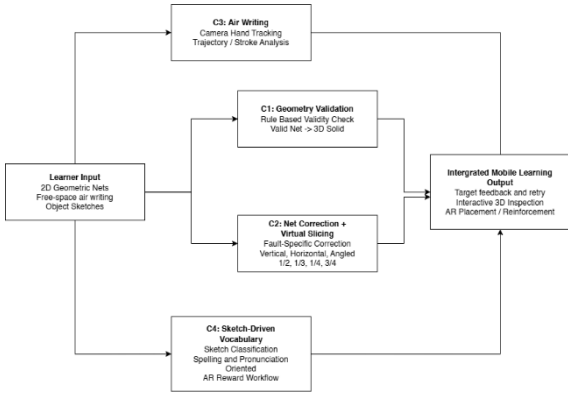


Fig.2. Integrated learner-input, analysis, feedback, and AR-reinforcement architecture.

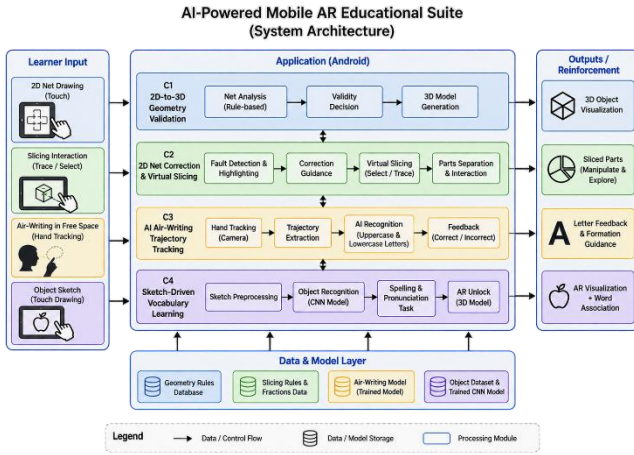


Fig.3. System Diagram

B. C1: 2D-to-3D Geometry Validation

C1 validates nets before 3D transformation so an invalid construction is not reinforced as a correct solid. Cube and cuboid checks use face connectivity and orientation propagation; the triangular prism requires three rectangular faces and two compatible triangular end faces; the tetrahedron checks four connected triangular faces and folding compatibility; and the cylinder checks two equal circular ends with a compatible rectangular lateral surface. Valid cases proceed to 3D generation and AR placement; invalid cases are routed to C2.

C. C2: 2D Net Correction and Virtual Slicing

For invalid nets, C2 identifies the fault location/relation, highlights it, gives a short child-oriented explanation, and shows the correct relation before retrying. The quantitative log available for this study focuses on the Virtual Slicing stage. A valid solid is placed on a detected plane; the learner selects Vertical, Horizontal, or Angled slicing and a fraction of 1/2, 1/3, 1/4, or 3/4, then traces the slicing guideline. A successful trace reveals separated prepared result meshes that can be inspected from different viewpoints.

D. C3: AI Air-Writing Trajectory Tracking

C3 tracks the learner's hand with the mobile camera and reconstructs the index-fingertip path while the learner writes

in free space. This preserves the temporal order of movement, allowing the component to reason about trajectory characteristics that a final raster image would hide. Recognition returns a predicted letter and confidence, while the interaction log preserves points, strokes, attempts, errors, and timing for evaluation. The term air-writing is used intentionally because the evaluated interaction is camera-based free-space writing, not screen drawing.

E. C4: Sketch-Driven Multisensory Vocabulary Learning

The learner draws a supported object, the CNN-based classifier predicts its class, and Unity loads the corresponding 3D representation. The recognized object becomes the context for spelling and pronunciation-oriented tasks. AR viewing is unlocked only when both tasks are completed successfully. This component is described as vocabulary learning rather than phonics because the evaluated workflow targets object-word association, spelling, spoken-word practice, and AR reinforcement rather than explicit grapheme-phoneme decoding.

V. RESULTS

A. Component-Level Technical and Workflow Results

C1 produced complete records for 25 learners. C1 was tested on five solids, a total of 25 cases, four valid cases and one invalid case per solid. The system classified all cases correctly: 20 valid and five invalid. All valid cases generated a 3D model and reached AR placement, whereas all invalid cases stopped before 3D/AR. Mean task time was 12.65 s (SD = 6.81 s). Mean time varied by shape: Cube 6.77 s, Cuboid 6.55 s, Triangular Prism 11.55 s, Cylinder 14.23 s, and Tetrahedron 24.14 s.

C2 recorded 25 Virtual Slicing sessions, of which 20 were successful (80.0%). Mean slicing duration was 6.40 s (SD = 1.91 s; range 3.02–9.78 s). Vertical slicing succeeded in 5/5 sessions, Horizontal in 11/14, and angled in 4/6. Fraction success was 4/5 for 1/2, 3/4 for 1/3, 6/7 for 1/4, and 7/9 for 3/4. These subgroup values are descriptive because the shape, direction, and fraction counts are unequal.

C3 contains 25 non-empty child-user air-writing trials. Nineteen were recognized correctly, three were incorrect predictions, and three failed to return a letter, giving 76.0% correct recognition when failed trials are counted as unsuccessful. Mean recognition confidence was 0.784 (SD = 0.314), mean drawing time was 5.05 s (SD = 2.07 s) and mean total session time was 8.35 s (SD = 2.65 s). The evaluated trials sampled 17 uppercase target letters; therefore, these results do not constitute a complete A–Z/a–z benchmark.

C4 contains 25 child-user sessions across Apple, Cat, Fish, Sun, and Tree. Spelling was completed in 20/25 sessions (80.0%), pronunciation-oriented acceptance was recorded in 21/25 (84.0%), and both conditions were satisfied in 17/25 (68.0%); exactly those 17 sessions unlocked and placed the AR model. Mean AI confidence was 0.832 (SD = 0.099), mean spelling attempts 1.40, mean pronunciation attempts 1.28 and mean total session time 56.52 s (SD = 13.22 s). Because the session log does not retain an independent intended-object ground-truth label, meaning confidence is not reported as classification accuracy.

TABLE II. TECHNICAL / WORKFLOW EVIDENCE

Component	n	Primary Metric	Observed Result
C1	25	Expected vs. system validity	25/25 agreement; 12.65 s mean
C2	25	Virtual Slicing completion	20/25 (80%); 6.40 s mean
C3	25	Air-written letter recognition	19/25 correct (76%); 8.35 s mean session time
C4	25	Full vocabulary/AR flow	17/25 (68%); 56.52 s mean session time

B. Learning Outcomes

The pre-post-test analysis showed the same absolute gain of four marks in every component. For learners aged 9–12, C1 and C2 each increased from an average of 5/10 to 9/10. For learners aged 6–8, C3 increased from 5/10 to 9/10 and C4 from 4/10 to 8/10. These are 40-percentage-point changes in the observed group means. Without participant-level paired scores in the final analysis file, the paper does not infer statistical significance from these aggregate values.

TABLE III. AGE-SPECIFIC PRE-/POST-TEST MEANS

Component	Age Group	Pre-Test Average (Out of 10)	Post-Test Average (Out of 10)	Change
C1: Geometry Validation	9-12	5	9	+4 Marks
C2: 2D Net Correction & Virtual Slicing	9-12	5	9	+4 Marks
C3: AI Air Writing Trajectory Tracking	6-8	5	9	+4 Marks
C4: Sketch Driven Multisensory Vocabulary Learning	6-8	4	8	+4 Marks

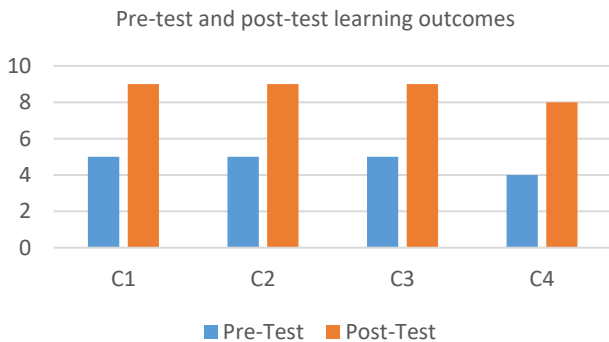


Fig.4. Pre-test and post-test learning outcomes

VI. DISCUSSION

RQ1 is supported by the stored C1 ground-truth labels: all 25 validation decisions matched the expected outcome, and only valid nets progressed to 3D/AR. This is stronger evidence than reporting the proportion of output labeled valid because it directly compares the decision with the expected state. RQ2 is supported for the quantitatively logged Virtual Slicing stage: 80% of sessions completed successfully, with more failures in angled than vertical slicing. The present log does not independently quantify fault-specific Net Correction, so the C2 contribution should be interpreted as a combined design contribution with quantitative evidence concentrated on slicing.

For RQ3, camera-based Air-Writing achieved 76% correct recognition in the 25 evaluated child-user trials. The incorrect and failed cases show why trajectory recognition for young learners remains more challenging than benchmark results reported for controlled air-writing datasets [13], [14]. The evaluated sample also covers only 17 uppercase targets, despite the implementation scope supporting upper- and lowercase letters; a complete alphabet benchmark remains necessary. RQ4 is answered at workflow level: 80% completed spelling, 84% received pronunciation-oriented acceptance, and 68% completed both conditions and reached AR placement. These measures demonstrate progression through the intended learning sequence but should not be interpreted as standalone speech-recognition or sketch-classification accuracy.

RQ5 is addressed by the pre-post-test means: each component showed a four-mark increase. The age split strengthens the interpretation because geometry and literacy were not evaluated as if a single 6–12 cohort had identical learning needs. RQ6 is addressed cautiously through observed completion, retrieved, and timing rather than a standardized usability score. The logs show that children could complete the intended workflows, but a dedicated validated child-usability instrument would be required for a stronger usability claim.

The integration rationale is also supported by the evaluation structure. C1–C4 do not share one model or one accuracy metric; instead, they share an educational interaction contract. This allows the suite to reuse navigation, feedback, 3D/AR presentation, and progression logic while preserving the technical soundness of domain-specific processing. The contribution is therefore not 'four features in one app,' but an architecture that consistently places computational interpretation between learner-generated input and digital reinforcement.

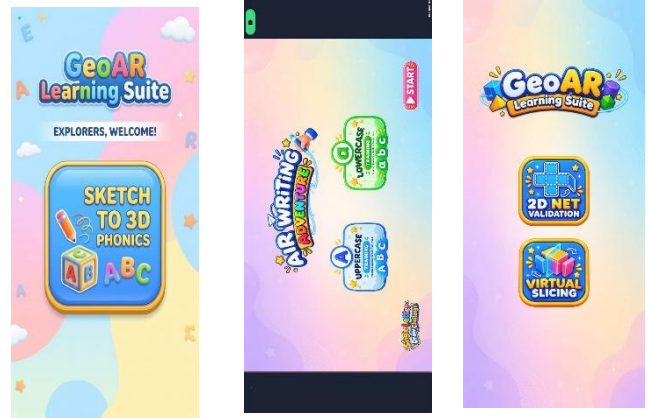


Fig.5. User Interface Designs

VII. LIMITATIONS AND FUTURE WORK

The study is bounded by a small, short-duration evaluation involving 50 unique learners: 25 in the geometry cohort and 25 in the literacy cohort. A randomized control condition or delayed retention test. The reported pre-/post-test evidence is based on aggregate means, so inferential statistics and effect sizes are not claimed. C2's final quantitative log covers Virtual Slicing but not a separate controlled dataset for fault-specific Net Correction. C3's child evaluation samples 17 uppercase targets rather than the complete upper-/lowercase scope. C4 user logs contain prediction confidence and downstream task states but not an independent intended-object label for every child sketch, preventing calculation of child-sketch classification accuracy from this file. Future work should expand participant numbers, retain participant-level paired scores, benchmark the complete C3 alphabet, add ground-truth labels for C4 child sketches, evaluate retention, and compare the integrated design against appropriate non-AR or predefined-input baselines.

VIII. CONCLUSION

This study presents and evaluates an Android AI/AR educational suite that unifies two curriculum domains through a common learner-centered architecture rather than a common algorithm. Learner-generated nets, free-space letter trajectories, and object sketches are interpreted before corrective or progression feedback, and 3D/AR reinforcement are shown. Across 50 unique learners divided into two age-specific cohorts, C1 achieved 25/25 validity agreement, C2 completed 20/25 slicing sessions, C3 correctly recognized 19/25 evaluated air-written letters, and C4 completed the full vocabulary-to-AR sequence in 17/25 sessions. Age-specific pre-/post-test means increased from 5/10 to 9/10 for C1–C3 and from 4/10 to 8/10 for C4. Within the stated limitations, the findings support the feasibility of using a shared input–analysis–feedback–reinforcement architecture for age-appropriate primary geometry and literacy learning while keeping component-level processing and evidence distinct.

IX. REFERENCES

- [1] H.-Y. Chang, T. Binali, J.-C. Liang, G.-L. Chiou, K.-H. Cheng, S. W.-Y. Lee, and C.-C. Tsai, "Ten years of augmented reality in education: A meta-analysis of (quasi-) experimental studies to investigate the impact," *Computers & Education*, vol. 191, Art. no. 104641, 2022, doi: 10.1016/j.compedu.2022.104641.
- [2] E. L.-C. Law and M. Heintz, "Augmented reality applications for K-12 education: A systematic review from the usability and user experience perspective," *International Journal of Child-Computer Interaction*, vol. 30, Art. no. 100321, 2021, doi: 10.1016/j.ijcci.2021.100321.
- [3] M. B. Nadzeri, M. Musa, C. M. Chew, and I. M. Ismail, "The effects of augmented reality geometry learning applications on spatial visualization ability for primary school pupils," *International Journal of Interactive Mobile Technologies*, vol. 18, no. 16, pp. 104–118, 2024, doi: 10.3991/ijim.v18i16.47079.
- [4] W. Tarnag, J.-K. Huang, and K.-L. Ou, "Improving elementary students' geometric understanding through augmented reality and its performance evaluation," *Systems*, vol. 12, no. 11, Art. no. 493, 2024, doi: 10.3390/systems12110493.
- [5] J. Wu, H. Jiang, L. Long, and X. Zhang, "Effects of AR mathematical picture books on primary school students' geometric thinking, cognitive load and flow experience," *Education and Information Technologies*, vol. 29, pp. 24627–24652, 2024, doi: 10.1007/s10639-024-12768-y.
- [6] A. Uriarte-Portillo, R. Zatarain-Cabada, M. L. Barrón-Estrada, M. B. Ibáñez, and L.-M. González-Barrón, "Intelligent augmented reality for learning geometry," *Information*, vol. 14, no. 4, Art. no. 245, 2023, doi: 10.3390/info14040245.
- [7] C. Walkington, M. J. Nathan, J. Washington, J. Hunnicutt, T. Darwin, L. Daugherty, and K. Schenck, "Comparing learning geometry using a tablet to head-mounted display augmented reality: How and when dimensionality matters," *Education and Information Technologies*, vol. 30, pp. 5397–5426, 2025, doi: 10.1007/s10639-024-13008-z.
- [8] S.-H. Zhang, Y.-C. Guo, and Q.-W. Gu, "Sketch2Model: View-aware 3D modeling from single free-hand sketches," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 6012–6021, doi: 10.1109/CVPR46437.2021.00595.
- [9] L. Zhang, "Hand-drawn sketch recognition with a double-channel convolutional neural network," *EURASIP Journal on Advances in Signal Processing*, vol. 2021, Art. no. 73, 2021, doi: 10.1186/s13634-021-00752-4.
- [10] A. K. Bhunia, P. N. Chowdhury, Y. Yang, T. M. Hospedales, T. Xiang, and Y.-Z. Song, "Vectorization and rasterization: Self-supervised learning for sketch and handwriting," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 5672–5681.
- [11] S. Mukherjee, S. A. Ahmed, D. P. Dogra, S. Kar, and P. P. Roy, "Fingertip detection and tracking for recognition of air-writing in videos," *Expert Systems with Applications*, vol. 136, pp. 217–229, 2019, doi: 10.1016/j.eswa.2019.06.034.
- [12] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.-L. Chang, and M. Grundmann, "MediaPipe Hands: On-device real-time hand tracking," in *CV4ARVR*, 2020.
- [13] M. S. Alam, K.-C. Kwon, M. A. Alam, M. Y. Abbass, S. M. Imtiaz, and N. Kim, "Trajectory-based air-writing recognition using deep neural network and depth sensor," *Sensors*, vol. 20, no. 2, Art. no. 376, 2020, doi: 10.3390/s20020376.
- [14] T. Watanabe, M. Maniruzzaman, M. A. M. Hasan, H.-S. Lee, S.-W. Jang, and J. Shin, "2D camera-based air-writing recognition using hand pose estimation and hybrid deep learning model," *Electronics*, vol. 12, no. 4, Art. no. 995, 2023, doi: 10.3390/electronics12040995.
- [15] A. Qedear, A. AlMatrafy, A. Al-Sowat, A. Saigh, and A. Alayed, "Real-time air-writing recognition for Arabic letters using deep learning," *Sensors*, vol. 24, no. 18, Art. no. 6098, 2024, doi: 10.3390/s24186098.
- [16] E. Korosidou, "The effects of augmented reality on very young learners' motivation and learning of the alphabet and vocabulary," *Digital*, vol. 4, no. 1, pp. 195–214, 2024, doi: 10.3390/digital4010010.
- [17] J. Belda-Medina and V. Marrahi-Gomez, "The impact of augmented reality (AR) on vocabulary acquisition and student motivation," *Electronics*, vol. 12, no. 3, Art. no. 749, 2023, doi: 10.3390/electronics12030749.
- [18] H. Lin and W. Wei, "A systematic review on vocabulary learning in AR and VR gamification context," *Computers & Education: X Reality*, vol. 4, Art. no. 100057, 2024, doi: 10.1016/j.cexr.2024.100057.
- [19] H. Weigand, P. Johannesson, and B. Andersson, "An artifact ontology for design science research," *Data & Knowledge Engineering*, vol. 133, Art. no. 101878, 2021, doi: 10.1016/j.datak.2021.101878.
- [20] S. G. Sutton, V. Arnold, P. Collier, and S. A. Leech, "Leveraging the synergies between design science and behavioral science research methods," *International Journal of Accounting Information Systems*, vol. 43, Art. no. 100536, 2021, doi: 10.1016/j.accinf.2021.100536.