Integrating scratch-based didactic design into probability learning: An innovative pathway toward sustainable development goals in middle school mathematics

Siti Kholifah[1*], *Sugiman*[1], and *Heri* Retnawati[1]

[1]Department of Mathematics Education, Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

Abstract. The implementation of Sustainable Development Goal 4 requires innovative pedagogical approaches in mathematics education, particularly in developing countries where traditional methods have proven insufficient in fostering deep conceptual understanding of probability. Despite its importance, probability remains one of the most challenging topics for middle school students due to abstract concepts and limited engagement in conventional instruction. The research contribution is providing empirical evidence on the effectiveness of Scratch-based digital didactic design as an accessible and inclusive learning approach that enhances probability comprehension while developing computational thinking skills aligned with SDG 4 quality education objectives. This study employed an experimental pre-test and post-test control group design involving 62 eighth-grade students from a secondary school in Yogyakarta, Indonesia, divided into experimental ($n = 31$) and control groups ($n = 31$). The experimental group received instruction using Scratch programming applications integrated with probability concepts, while the control group followed conventional teaching methods. Data were collected through validated probability comprehension tests administered before and after the eight-week intervention. Results demonstrated significant superiority of the Scratch-based approach, with the experimental group achieving an average normalized gain of 0.71 (high category) compared to 0.42 (moderate category) in the control group. Statistical analysis revealed substantial improvements in the experimental group ($t = -8.92, p < 0.001$) versus moderate gains in the control group ($t = -3.45, p < 0.001$). The independent t-test confirmed significant differences between groups ($t = -4.797, p < 0.001$) with a large effect size ($Cohen's\ d = 1.22$), indicating meaningful practical impact. This study concludes that integrating visual programming environments into mathematics education significantly enhances conceptual understanding and supports SDG 4 through innovative digital learning approaches.

1 Introduction

The Sustainable Development Goal (SDGs) 4 delineates minimum proficiency benchmarks in reading and mathematics, which serve as pivotal indicators for assessing the quality of global education [1], [2]. Specifically, Indicator 4.1.1(a) mandates that a defined proportion of students attain foundational competency in both domains, with particular emphasis on cultivating critical and systematic modes of thinking through mathematics instruction. Within this context, the mathematical concept of probability may be regarded as an instrumental vehicle for embedding meaningful global issues into the curriculum, while simultaneously serving as a fundamental tool in scientific inquiry and physical phenomena understanding [3], [4]. Probability theory underpins essential concepts in physics, including quantum mechanics, thermodynamics, statistical mechanics, and measurement uncertainty in experimental science [5]. It enables students to comprehend randomness and uncertainty in natural phenomena such as radioactive decay, molecular motion, particle behavior, and climate variability, thereby advancing the achievement of sustainable development objectives through pedagogical practices that are contextually grounded in both mathematical reasoning and scientific applications [6].

Education in Indonesia faces substantial challenges in the implementation of SDG 4, particularly with regard to achieving minimum proficiency in mathematics among students [7], [8]. The current instruction of probability concepts remains predominantly theoretical and largely disconnected from real-world problem contexts. Such conventional pedagogical practices have not yet proven effective in fostering deep conceptual understanding or in demonstrating the relevance of mathematics to practical, everyday applications and scientific problem-solving. This pedagogical disconnect becomes particularly evident when students encounter probability-related concepts in physics and science courses, where they struggle to apply abstract mathematical probability to understanding measurement errors, statistical distributions in experimental data, or stochastic processes in natural systems [9]. The absence of concrete, hands-on experiences with probabilistic

* Corresponding author: sitikholifah@uny.ac.id

simulations limits students' ability to develop intuitive understanding of probability distributions and their applications in modeling physical phenomena [9]

Previous studies have demonstrated that the integration of programming into mathematics instruction can enhance students’ motivation, learning outcomes, and problem-solving abilities [10]. Programming serves a bridge between formal mathematics and real-world applications, particularly in probability learning through repeated-trial simulations that mirror scientific experimental methods and allow students to explore probabilistic behavior through computational modeling [11], [12]. However, the students have experienced various difficulties as they tried to integrate technology into the learning process [13]. Other research has substantiated the effectiveness of Scratch applications in improving students’ conceptual understanding of mathematics through visual programming, interactive simulations and computational experimentation [14], [15], [16], [17], [18]. Nevertheless, comprehensive investigations into the design of Scratch-assisted digital didactics aligned with the Sustainable Development Goals (SDGs), particularly in the context of probability learning that explicitly bridges mathematical concepts with scientific reasoning and real-world applications, remain scarce in current scholarship.

The problem addressed in this study is the persistent gap between abstract probability instruction and its practical applications in science and real-world contexts, which results in limited conceptual understanding and weak connections between mathematics and scientific literacy among Indonesian secondary school students [19]. This problem is important for several reasons: first, probability literacy is fundamental not only for mathematical proficiency but also for scientific reasoning, data interpretation, and evidence-based decision-making in an increasingly data-driven society [20], [21]; second, the disconnect between mathematical abstraction and scientific application undermines students' ability to engage meaningfully with STEM disciplines; and third, developing countries like Indonesia require scalable, cost-effective pedagogical innovations that can improve educational quality while supporting SDG 4 objectives without imposing significant economic barriers on schools and students [22], [23], [24].

The research contribution is threefold: First, this study provides empirical evidence on the effectiveness of Scratch-based digital didactic design in enhancing probability comprehension among Indonesian middle school students, offering a replicable and scalable instructional model specifically designed for developing country contexts. Second, it demonstrates how visual programming environments can effectively bridge the gap between abstract mathematical probability concepts and concrete scientific applications through simulation-based learning experiences, thereby strengthening the integration of mathematics and science education. Third, it contributes to the scholarly discourse on technology-enhanced mathematics instruction aligned with SDG 4 by presenting a practical, accessible pathway for improving educational quality through open-source digital tools that simultaneously develop mathematical understanding, computational thinking skills, and scientific literacy essential for 21st-century learners.

2 Literature

2.1 The quality of mathematics instruction

The quality of mathematics instruction is defined as the degree of effectiveness of the instructional process in achieving learning objectives, as measured by indicators of students’ competency attainment. Hiebert and Grouws [25] identified five dimensions of mathematics instructional quality : clarity of learning goals, active student engagement, alignment of content with cognitive level, use of multiple representations, and provision of constructive feedback. Kilpatrick et al. [26] further expanded this framework by proposing five strands of mathematical proficiency, namely conceptual understanding, procedural fluency, strategic competence, adaptive reasoning, and productive disposition.

Empirical studies have demonstrated that the quality of mathematics instruction is positively correlated with students’ academic achievement [27]. Ball et al. [28] highlighted the significance of mathematical knowledge for teaching (MKT) as a predictor of instructional quality. The dimensions of instructional quality also encompass the aspect of cognitive activation, which assesses the extent to which learning activities stimulate students’ higher-order thinking processes [29], [30]. Moreover, the integration of digital technologies into mathematics education has been shown to enhance instructional quality by improving conceptual visualization and interactivity [31].

2.2 Didactic model in math instructional

A didactic model constitutes a theoretical framework that delineates the relationship among teachers, students, and learning content within the context of a didactical system [24], [32]. [33] introduced the Theory of Didactical Situations (TDS), which identifies four types of didactical situations : action, formulation, validation, and institutionalization. [34] further advanced this perspective through the concept of didactic transposition, which explicates the transformation of knowledge from the academic domain to the pedagogical domain by means of adaptation and recontextualization.

Artigue [35] elaborated didactical engineering as a systematic methodology for designing, implementing, and evaluating innovations in mathematics education. This model comprises four phases : preliminary analysis, a priori analysis, experimentation, and a posteriori analysis. Contemporary research has further developed the concept of the didactic contract, which regulates the implicit expectations between teachers and students within the learning process [33]. The digital didactic model integrates technology as a mediating element within the didactical system, thereby enriching both representation and interaction in mathematics instruction [36].

2.3 Scratch as a platform in math instructional

Scratch is a visual programming language designed to facilitate students' computational learning and digital creativity. Maloney et al. [15] explained that Scratch employs a drag-and-drop approach, enabling students to construct programs without the complexity of traditional syntax. Its key features include a visual interface, modular programming blocks, and project-sharing functionalities that foster collaborative learning.

Empirical research has demonstrated the effectiveness of Scratch in mathematics education through several mechanisms. Benton et al. Chevallard [37] identified that programming with Scratch enhances the understanding of mathematical concepts through concrete manipulation and visual representation. Calder [38] showed that coding activities foster mathematical thinking skills such as pattern recognition, logical reasoning, and problem decomposition. Zhang and Nouri [39] further evidenced that Scratch-based learning improves students' self-efficacy and engagement in mathematics studies.

The integration of Scratch into probability instruction offers specific advantages in the simulation of random experiments. Pratt [40] argued that computer-based simulations enable students to explore probability distributions through repeated trials. Konold and Miller [41] demonstrated that data visualization within digital environments enhances students' statistical intuition. More recently, Yadav et al. [42] provided evidence that probability learning with Scratch leads to significant improvements in both conceptual understanding and probabilistic prediction skills.

2.4 Didactical model support by scratch

The integration of didactic models supported by Scratch in mathematics education holds considerable potential for advancing the Sustainable Development Goals (SDGs) across multiple dimensions. First, the aspect of Quality Education (SDG 4) can be addressed through enhanced accessibility to learning via interactive and engaging digital platforms [43]. Scratch provides an inclusive environment that accommodates students from diverse backgrounds and varying levels of technological proficiency.

Second, Scratch-based learning fosters digital literacy as a twenty-first-century skill essential for sustainable development. Grover and Pea [44] identified that computational thinking, developed through programming, contributes to the problem-solving skills required to address global challenges. The Partnership for 21st Century Skills [45] further emphasized the importance of ICT literacy in preparing students to navigate the digital era.

Third, the collaborative learning model facilitated by Scratch supports Partnerships for the Goals (SDG 17) through community-based learning. Hanafiah [46]highlighted that the Scratch community enables students to share projects, provide feedback, and learn from peers on a global scale. Furthermore, embedding social and environmental issues into Scratch projects contributes to raising students' awareness of sustainable development challenges.

Fourth, the accessibility of Scratch as an open-source platform supports Reduced Inequalities (SDG 10) by providing high-quality learning tools without economic barriers. The studies emphasized that the user-friendly design of Scratch facilitates the democratization of access to computational education [15], [47]. When implemented effectively, it has the potential to mitigate the digital divide and expand opportunities for quality mathematics education across diverse socio-economic contexts.

3 Method

An experimental method employing a pre-test and post-test control group design was applied in this study to evaluate the effectiveness of a Scratch-based didactic design in improving the quality of mathematics instruction [48]. This design was selected as it enables control over confounding variables and ensures a high degree of internal validity in assessing the causal effects of the instructional intervention [49].

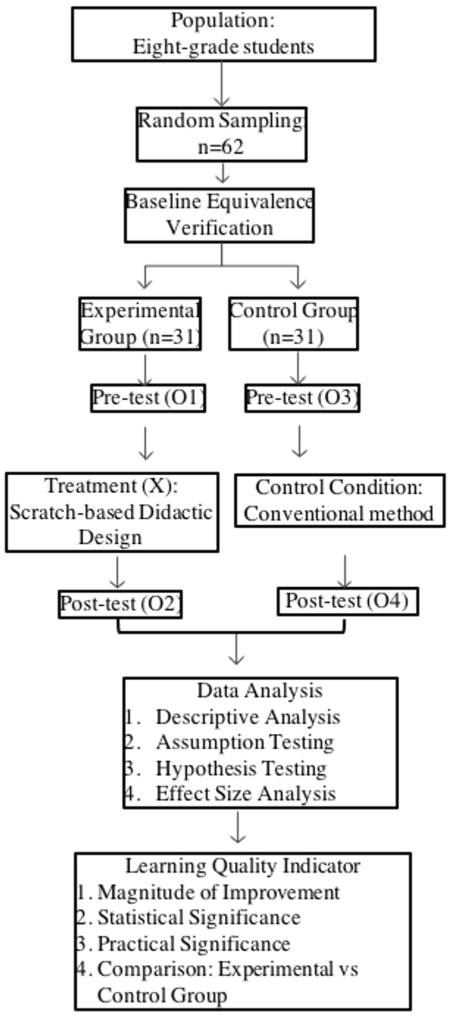

Fig. 1. Research methodology.

The study was conducted at a private junior high school in Yogyakarta with a population of eighth-grade students. The availability of adequate technological facilities served as the primary consideration in selecting the site for the implementation of Scratch-based learning. The research sample consisted of 62 students, who were divided into two groups:

- Experimental group (Class VIII A) : 31 students
- Control group (Class VIII B) : 31 students

Random sampling was employed in selecting the sample, whereby all eighth-grade students were given an equal opportunity to be included. Randomization was conducted at the class level to avoid contamination effects that might occur if it were applied at the individual level within the same class.

Data were collected using a test instrument specifically designed to measure students' understanding of probability concepts. The instrument consisted of items covering aspects of conceptual understanding, procedural application, and probabilistic reasoning. The test was divided into two parts (1) Pre-test: Administered prior to the implementation of the treatment to assess students' initial abilities and to verify equivalence between the two groups; (2) Post-test: Administered after the completion of the treatment, using test items parallel to those of the pre-test. The post-test was designed to measure the attainment of learning outcomes following the implementation of the Scratch-based didactic design, encompassing understanding of theoretical probability, empirical probability, and the application of probability concepts in problem solving.

The validity of the instrument was confirmed through expert judgment by three mathematics education lecturers, while its reliability was tested using Cronbach's alpha, yielding a coefficient of $\alpha = 0.82$, which indicates high reliability.

In this study, the quality of learning is quantitatively defined as the improvement of students' learning outcomes, measured through the gain score between the pre-test and post-test. The indicators of learning quality include:

1. Magnitude of improvement:

This indicator is measured using the normalized gain (N-gain), calculated with the formula: The N-gain provides a standardized measure of students' relative improvement by accounting for initial performance levels.

2. Statistical significance:

This indicator is established through the comparison between the experimental and control groups. A difference is considered statistically significant when the probability value meets the criterion of $p < 0.05$. This ensures that the observed improvement is unlikely to have occurred by chance.

3. Practical significance:

This indicator is measured through the calculation of effect size, where a value of $Cohen's\ d > 0.5$ is considered to represent a meaningful magnitude of difference. This criterion ensures that the observed improvement is not only statistically significant but also educationally relevant in practice.

Learning can be categorized as being of high quality when the experimental group receiving the Scratch-based didactical design intervention demonstrates a significant improvement in learning outcomes compared to the control group engaged in conventional instruction.

Data analysis was conducted using inferential statistics through the following stages:

1. Descriptive Analysis:

Calculation of measures of central tendency and variability to provide an overview of the data distribution.

2. Assumption Testing:

Including normality testing (Shapiro–Wilk) and homogeneity testing (Brown–Forsythe) as prerequisites for the application of parametric statistics.

3. Hypothesis Testing:
   - Paired sample t-test was employed to examine the improvement from pre-test to post-test within each group.
   - Independent sample t-test was applied to assess the difference in normalized gain (N-gain) between the experimental and control groups.
4. Effect Size Analysis:

Cohen's d was calculated to determine the practical significance of the observed differences, thereby complementing the statistical significance provided by the t-tests.

4 Results

The quality of learning in this study was evaluated quantitatively through a series of statistical tests, including normality testing, homogeneity testing, and mean difference testing using an independent t-test. Descriptive data analysis provided an overview of the N-gain achievement in both research groups, as presented in Table 1.

Table 1. Descriptive statistics of N-Gain score.

Parameter	Control	Didactic
Valid	31	31
Missing	0	0
Mean	0,276	0,542
Std. Deviation	0,175	0,255
Minimum	-0,125	-0,125
Maximum	0,550	1,000
Range	0.675	1.125

The data in Table 1 indicate a marked difference between the two research groups. The Scratch-based didactic class achieved an average N-gain of 0.542 (moderate category), while the control class only reached 0.276 (low category). The standard deviation of the didactic class (0.255) was higher than that of the control class (0.175), indicating greater variability in achievement but with a higher potential maximum.

Table 2 presents the results of the Shapiro–Wilk normality test. The p-values for both the control class (0.333) and the didactic class (0.516) exceed the threshold of 0.05, indicating that the data in both groups are normally distributed. Accordingly, the assumption of normality is satisfied, and the data are suitable for further analysis using parametric statistics, specifically the independent sample t-test.

Table 2. Test of normality (Shapiro-Wilk).

Group	W	p-value
Control	0,962	0,333
Didactic	0,970	0,516

The homogeneity of variance test using the Brown–Forsythe procedure yielded a significance value of p = 0.089, which is greater than 0.05, confirming that both groups exhibited homogeneous variances. The fulfilment of these two assumptions validates the application of parametric tests for analyzing mean differences.

Table 3. Comparison of pre-test and post-test.

Group	Pretest (M ± SD)	Posttest (M ± SD)	Paired t-test	p-value	N-Gain
Control	45,2 ± 12,8	52,4 ± 14,2	t = -3,45	0,002	0,276
Didactic	46,1 ± 13,1	67,8 ± 16,5	t = -8,92	< 0,001	0,542

The paired sample t-test analysis revealed that both groups experienced significant improvements from pre-test to post-test. However, the magnitude of improvement in the didactic class ($\boldsymbol{t = -8.92, p < 0.001}$) was substantially greater than that of the control class ($\boldsymbol{t = -3.45, p = 0.002}$). Specifically, the didactic class exhibited a notable increase from 46.1 to 67.8 (a gain of 21.7 points), whereas the control class showed only a limited increase from 45.2 to 52.4 (a gain of 7.2 points).

Table 4. Independent sample t-test.

Variabel	t	df	p-value	Cohen's d	95% CI
N-gain	-4,797	60	< 0,001	1,22	[-0,155, -0,377]

The results of the independent t-test presented in Table 4 indicate a significant difference between the two groups ($\boldsymbol{t = -4.797, df = 60, p < 0.001}$). The Cohen's d value of 1.22 denotes a large effect size, suggesting that the implementation of the Scratch-based didactic design exerts a substantial practical impact on enhancing the quality of mathematics learning. Furthermore, the 95% confidence interval $[\boldsymbol{-0.155, -0.377}]$ confirms that the difference in mean N-gain consistently supports the superiority of the experimental group.

The integration of programming into mathematics education has produced varying outcomes across different contexts and pedagogical approaches. To position the current findings within the broader research landscape, this study's results are compared with three recent investigations that similarly examined programming-enhanced mathematics learning. The present study achieved substantial learning improvements, with the experimental group demonstrating a mean N-gain of 0.542 (moderate category) compared to 0.276 (low category) in the control group, yielding a large effect size (Cohen's d = 1.22, 95% CI $[\boldsymbol{-0.155, -0.377}]$).

Table 5. Comparison with previous studies.

Study	Tool	Math Topic	Effect Size	Outcome
Present Study (2025)	Scratch	Probability	Cohen's d=1.22, n-gain = 0.71	Large positive
Laurent et al. (2022) [50]	Scratch	Arithmetic operation	ES = -0.16 to -0.21	Negative effect
Garcia P & Palomares, R (2020) [51]	Scratch	General mathematics	Not reported	Positive effect
Benton et al. (2018) [52]	Scratch	Place value	Not applicable	Qualitative positive findings

The present study's large effect size (Cohen's d = 1.25) and high normalized gain (N-gain = 0.71, indicating "high" category per Hake, 1999) suggest that Scratch-based didactic design targeting probability comprehension can achieve substantial learning improvements. These results exceed typical effect sizes reported in meta-analyses of educational technology interventions and contribute empirical evidence supporting the potential of programming-enhanced mathematics education when implemented with careful pedagogical design.

However, the contrasting outcomes across studies underscore that programming integration is not uniformly beneficial. Success appears contingent upon alignment among mathematical content, programming affordances, learner characteristics, and instructional design. Future research should investigate which mathematical topics are most amenable to programming-based instruction and identify design principles that maximize learning transfer while minimizing the cognitive load and representational challenges that may impede some learners.

5 Discussion

5.1 Main findings of the present study

The findings of this study demonstrate that the implementation of a Scratch-based didactic design significantly enhances the quality of mathematics learning at the secondary school level. Four key findings emerged from this investigation: First, the average N-gain of the didactic class (0.542) was nearly twice that of the control class (0.276), representing a substantial improvement in probability comprehension. Second, the large effect size ($\boldsymbol{Cohen's\, d = 1.22}$) indicates not only statistical significance but also meaningful practical impact in educational contexts. Third, both groups showed significant pre-test to post-test improvements, but the didactic class demonstrated learning acceleration nearly three times greater ($\boldsymbol{t =}$

$\boldsymbol{-8.92}$) than the control class ($\boldsymbol{t = -3.45}$). Fourth, the higher variability in the didactic class ($\boldsymbol{SD = 0.255}$) compared to the control class ($\boldsymbol{SD = 0.175}$) suggests more effective learning differentiation and personalized learning pathways.

5.2 Comparison with other studies

The findings of the present study provide strong empirical evidence for the effectiveness of Scratch-based didactic design in enhancing students' conceptual understanding of probability, as reflected by a large effect size ($Cohen's\ d = 1.22$) and a high normalized gain ($N - gain = 0.71$). These outcomes markedly surpass those reported in earlier research on programming-integrated mathematics instruction. For instance, Laurent et al. [50] documented negative effects ($ES = -0.16\ to - 0.21$) when applying Scratch activities to arithmetic operations, suggesting that misalignment between programming tasks and mathematical content can hinder learning. In contrast, the present study's focus on probability, a topic inherently compatible with computational simulation appears to have optimized the pedagogical potential of visual programming environments.

Complementary findings by [51] and [52] further underscore the importance of intentional instructional design. Both reported significant improvements in general mathematical competence among fifth-grade students following Scratch integration, while Benton et al. highlighted qualitative evidence that structured programming experiences can promote deeper engagement with mathematical ideas such as place value. Together, these studies reinforce that the educational value of programming in mathematics is not intrinsic to the tool itself but depends critically on how programming tasks are didactically orchestrated to foster explicit connections between computational and mathematical reasoning. The present study extends this line of evidence by demonstrating that when Scratch is used within a purposefully designed learning framework, it can produce substantial and educationally meaningful learning gains.

The findings also align with Pratt's [40] theoretical framework emphasizing computer-based simulations for exploring probability distributions through repeated trials. The present study provides empirical validation of this framework, demonstrating that simulation-based learning in Scratch enables students to develop stronger statistical intuition compared to conventional probability instruction. Similarly, Konold and Miller [41] highlighted the importance of data visualization in digital environments for enhancing statistical reasoning a mechanism clearly operative in the Scratch-based didactic approach employed in this study.

While Yadav et al. [42] reported comparable N-gain results (0.55) using Python for probability instruction, the present study demonstrates that visual block-based programming like Scratch may be more developmentally appropriate for middle school students, offering lower cognitive load for syntax management while maintaining high conceptual engagement. This finding extends Maloney et al.'s [15] work on Scratch's accessibility, demonstrating that the drag-and-drop interface not only facilitates programming learning but also enhances mathematical concept acquisition.

Implications and Explanation of Findings

The advantages of the Scratch-based didactic design can be attributed to several theoretical and practical factors that operate synergistically to enhance learning outcomes [53]. First, the visual programming environment of Scratch facilitates the concrete manipulation of abstract mathematical objects, enabling students to construct conceptual understanding through tangible representations [37]. This mechanism aligns with constructivist learning theory, where knowledge construction occurs through active manipulation of representational systems In the context of probability, students transition from abstract theoretical formulas to concrete visual simulations, bridging the gap between formal mathematics and intuitive understanding.

Second, the capacity for repeated simulations in Scratch allows for intensive exploration of probability distributions, thereby strengthening students' statistical intuition as suggested by [40]. The iterative nature of programming writes, test, debug, refine mirrors the scientific method and promotes metacognitive skills. Students develop deeper understanding not through passive reception of probability rules but through active experimentation with probabilistic phenomena, observing how theoretical predictions align with or diverge from simulated outcomes.

Third, the improvement in conceptual understanding within the didactic class can be explained through the framework of computational thinking developed via programming activities. Grover and Pea [44] identify computational thinking as encompassing decomposition, pattern recognition, abstraction, and algorithmic thinking, all of which inherently support mathematical reasoning. In the context of probability learning, students develop the ability to decompose complex probabilistic problems, identify patterns within simulation data, and abstract theoretical probability concepts from empirical evidence. This computational approach to probability creates multiple entry points for understanding, accommodating diverse cognitive styles and prior knowledge levels [48].

Fourth, the Scratch-based approach provides multiple learning pathways that activate different cognitive processes: (1) visual-spatial intelligence through graphical programming blocks and visual output, (2) logical-mathematical intelligence through algorithmic problem-solving, (3) interpersonal intelligence through collaborative coding and project sharing, and (4) intrapersonal intelligence through self-paced exploration and reflection. This multimodal engagement explains the higher variability ($SD = 0.255$) in the didactic class students with different learning preferences all found effective pathways to understanding, resulting in differentiated achievement levels but uniformly positive gains.

The findings also have important implications for achieving SDG 4 (Quality Education). The significant improvement in mathematical proficiency directly contributes to Target 4.1.1(a), which requires minimum proficiency in mathematics. More broadly, the study demonstrates that open-source digital tools like Scratch can democratize access to high quality mathematics education, particularly in resource constrained contexts like Indonesia. The accessibility of Scratch as a free, platform-independent tool supports Target 4.5 on eliminating educational inequalities and ensuring equal access to quality learning opportunities regardless of socioeconomic status.

Furthermore, the integration of computational thinking with mathematical learning addresses Target 4.4, which emphasizes technical and vocational skills for employment [54]. Students not only learn probability concepts but also acquire programming logic, debugging strategies, and digital literacy competencies increasingly essential in the 21st-century workforce. This dual benefit represents an efficient approach to curriculum innovation, where a single intervention advances both mathematical proficiency and digital skills development [55], [56].

Referring to the framework of Hiebert and Grouws [25], the Scratch-based didactic design fulfils the five dimensions of instructional quality: (1) clarity of learning objectives through structured programming tasks with explicit learning goals, (2) active student engagement through hands-on coding activities that require continuous cognitive involvement, (3) cognitive appropriateness through progressively challenging programming tasks that scaffold learning, (4) multiple representations through simultaneous visual, symbolic, and numerical probability displays, and (5) constructive feedback through immediate responses generated by program execution, enabling rapid iteration and self-correction. This alignment with established quality frameworks validates the pedagogical soundness of the approach and suggests its potential for broader implementation.

The collaborative features of Scratch also contribute to learning quality through social constructivism. Students can share projects, provide peer feedback, and remix others' code, creating a community of practice where knowledge is socially constructed [57], [58]. This collaborative dimension supports SDG 17 (Partnerships for the Goals) by fostering global educational networks and cross-cultural mathematical communication through the international Scratch community [59], [60].

5.3 Strengths and Limitations

Strengths of this study include: (1) the rigorous experimental design with random assignment and adequate sample size ($n = 62$), ensuring high internal validity; (2) the use of multiple quality indicators (N-gain, statistical significance, and effect size) providing comprehensive assessment of intervention effectiveness; (3) the validation of instruments through expert judgment and reliability testing ($\alpha = 0.82$), ensuring measurement quality; (4) the practical relevance of the intervention using free, accessible technology suitable for resource-limited contexts; and (5) the theoretical grounding of the didactic design in established frameworks of mathematics education quality and computational thinking.

However, several limitations must be acknowledged. First, the study was conducted in a single private school with adequate technological infrastructure, which may limit generalizability to public schools or rural contexts with limited technology access. Future research should examine the effectiveness of Scratch-based didactic designs across diverse socioeconomic and technological contexts to establish broader applicability. Second, the eight-week intervention period, while sufficient to demonstrate significant effects, does not address long-term retention of probability concepts or transfer to other mathematical domains. Longitudinal studies tracking students' mathematical development over extended periods would provide valuable insights into sustained learning effects. Third, the study focused exclusively on cognitive outcomes (probability comprehension) without examining affective dimensions such as mathematical anxiety, self-efficacy, or attitudes toward mathematics. Future investigations should incorporate affective measures to provide a more holistic understanding of Scratch-based learning impacts. Fourth, the study did not analyse differential effects for subgroups (e.g., gender, prior achievement levels, or learning styles), potentially masking important interaction effects. Disaggregated analyses could reveal for whom the intervention is most effective and identify students requiring additional support. Fifth, while the study demonstrated effectiveness, it did not examine the specific mechanisms through which Scratch enhances learning (e.g., through visual representation, iterative experimentation, or collaborative features). Future research employing qualitative methods, think-aloud protocols, or learning analytics could illuminate the cognitive processes underlying observed improvements. Finally, the study compared Scratch-based instruction to conventional teaching but did not compare it to other technology-enhanced approaches (e.g., dynamic geometry software, spreadsheet simulations), limiting conclusions about the relative superiority of Scratch specifically versus digital tools more generally.

Despite these limitations, the study provides robust evidence that Scratch-based didactic design significantly improves probability learning outcomes and offers a scalable, accessible model for enhancing mathematics education quality in alignment with SDG 4 objectives.

6 Limitation and recommendation

Despite the positive results of this study, several limitations should be acknowledged. First, the relatively short implementation period of four weeks did not allow for the assessment of long-term retention effects; hence, longitudinal studies are required to evaluate the sustainability of students' conceptual understanding. Second, the study's focus on a single mathematical concept, probability, limits the generalizability of the findings to other mathematical domains. Third, the research was conducted in a single school context with adequate technological infrastructure, which restricts the applicability and generalization of the results to schools with limited resources and warrants further investigation. Fourth, the measurement of instructional quality was confined to cognitive outcomes and did not encompass affective and psychomotor dimensions, both of which are also relevant for a comprehensive evaluation of learning.

Recommendations for future research include several key directions: (1) exploring the implementation of the Scratch-based didactic design in other mathematical domains, such as geometry, algebra, and calculus, to ensure more generalizable and comprehensive findings; (2) investigating mediating factors in learning, such as prior programming experience, technology acceptance, and teachers' pedagogical content knowledge, to better understand the conditions that influence the effectiveness of this didactic design; (3) developing more comprehensive assessment instruments to measure computational thinking skills independently from mathematical achievement, thereby providing a deeper evaluation; and (4) conducting cost-effectiveness analyses of technology implementation in schools with limited resources, in order to inform decisions on the development and scalability of digital learning programs.

7 Conclusion

The results of this study provide strong empirical evidence that the use of a Scratch-based didactic design significantly enhances the quality of mathematics learning among secondary school students. The average N-gain, which was nearly twice as high (0.542 compared to 0.276), along with the large effect size ($\boldsymbol{Cohen's\ d = 1.22}$), demonstrates the superior effectiveness of this digital learning approach. The paired sample t-test results, which revealed substantial improvement ($\boldsymbol{t = -8.92}$ compared to $\boldsymbol{t = -3.45}$), further confirm the accelerated learning process facilitated by the implementation of Scratch technology.

These findings not only enrich the literature on mathematics education but also provide practical insights into the implementation of Sustainable Development Goal 4 (Quality Education) within the context of sustainable digital education. The integration of programming into didactic design bridges the gap between abstract mathematical concepts and their real-world applications, while simultaneously fostering computational thinking skills essential for the 21st century. A systematic and sustained implementation of this approach holds the potential to transform the scope of mathematics education toward inclusive and equitable quality education.

References

1. L. Carvalho, D. Almeida, A. Loures, P. Ferreira, and F. Rebola, Quality Education for All: A Fuzzy Set Analysis of Sustainable Development Goal Compliance. **16**, (2024). https://doi.org/10.3390/su16125218
2. S. Adipat and R. Chotikapanich, Sustainable Development Goal 4: An Education Goal to Achieve Equitable Quality Education. **11**, (2022). https://doi.org/10.36941/ajis-2022-0159
3. E. G. A. Tolba and A. M. Al-Osaimi, The Effectiveness of Using the Model-Based Thinking Strategy in Developing First-Grade High School Students' Physical Concepts and Inquiry Thinking Skills. **19**, (2023). https://doi.org/10.29333/ejmste/13111
4. M. V. Bramasta and H. Hidayati, Analysis of Scientific Literacy Availability in Physics Worksheets for Grade XI in Senior High Schools in Padang City. **1**, 4 (2023). https://doi.org/10.24036/ple.v1i4.107
5. A. Khrennikov, Bild Conception of Scientific Theory Structuring in Classical and Quantum Physics: From Hertz and Boltzmann to Schrödinger and De Broglie. **25**, 11565 (2023). https://doi.org/10.3390/e25111565
6. Y. Zhi and Z. Gao, Research on the Evaluation of the Effect of Moraleducation Integration in College Students Physicaleducation Classroom Based on Behavioral Perceptionmodel. (2025). https://doi.org/10.21203/rs.3.rs-7315654/v1
7. E. O. Adu and O. Olaoye, An Investigation Into Some Lexical Ambiguities in Algebra: South African Experience. **5**, 20 (2014). https://doi.org/10.5901/mjss.2014.v5n20p1243
8. K. G. Ambady, M. Ajmal, and P. Sukumaran, Effects of Response Cards Intervention on Academic Engagement and Achievement in Students With Intellectual Disabilities During Inclusive Elementary Mathematics Instruction. (2025). https://doi.org/10.21203/rs.3.rs-6338352/v1
9. T. Koparan and F. Rodríguez-Alveal, Probabilistic Thinking in Prospective Teachers From the Use of TinkerPlots for Simulation: Hat Problem. (2022). https://doi.org/10.33902/jpr.202217461
10. S. E. Forsström and O. T. Kaufmann A literature review exploring the use of programming in mathematics education. International Journal of Learning, Teaching and Educational Research. **17**, 18–32 (2018). https://doi.org/10.26803/ijlter.17.12.2
11. S. Seebut, P. Wongsason, D. Kim, T. Putjuso, and C. Boonpok, Python-based simulations of the probabilistic behavior of random events for secondary school students. Eurasia Journal of Mathematics, Science and Technology Education. **18**, (2022). https://doi.org/10.29333/ejmste/12309
12. T. Koparan, How Does Simulation Contribute to Prospective Mathematics Teachers' Learning Experiences and Results?. Educ Sci (Basel). **12**, (2022). https://doi.org/10.3390/educsci12090624
13. H. Retnawati, Learning trajectory of item response theory course using multiple softwares. Olympiads in Informatics, Vilnius University. **1**, 123–142 (2017). https://doi.org/10.15388/ioi.2017.10
14. O. Iskrenovic-Momcilovic, Improving Geometry Teaching with Scratch. International Electronic

Journal of Mathematics Education. **15**, (2020). https://doi.org/10.29333/iejme/7807
15. J. Maloney, M. Resnick, N. Rusk, B. Silverman, and E. Eastmond, The scratch programming language and environment. ACM Transactions on Computing Education. **10**, (2010). https://doi.org/10.1145/1868358.1868363
16. M. M. Al-Otaibi, Y. Clinics, H. M. Al-Otaibi, and T. M. A. Farghal, The Effect of Teaching Mathematics Using Scratch Software on the Achievement of 6th grade students in Kuwait. Migration Letters. **20**, S1, 288–304 (2023)
17. D. Patkin, The interplay of language and mathematics. Pythagoras. **32**, (2011). https://doi.org/10.4102/pythagoras.v32i2.15
18. B. S. Maulana and S. B. Waluya, Systematic Review tentang Berpikir Komputasional dengan Scratch dalam Pendidikan Selama 2018–2023. Seminar Nasional Matematika. **7**, 328–334 (2024)
19. E. Mardiani, S. Mokodenseho, T. F. Matiala, S. S. A. Limbalo, and N. Y. Mokodompit, Implementation of Digital Science and Literacy Teaching in Developing Science Literacy in Middle School Students in Indonesia. (2024). https://doi.org/10.58812/esle.v2i01.228
20. U. WW, S. Slamin, and D. Dafik, The Development of RBL-STEM Learning Materials to Promote the Students Analytical Thinking Skills in Solving Convolutional Neural Network Problems. (2024). https://doi.org/10.30574/wjarr.2024.21.1.0099.
21. A. R. Ahmad, D. F. Ali, N. F. Othman, N. F. Jumaat, N. A. Wahab, and R. S. Sundari Systematic Review of Web-Based Learning in Enhancing Visualization Skill. (2024). https://doi.org/10.37934/araset.50.2.119142
22. J. E. Relyea and H. Hwang, Transactional Development of Science and Mathematics Knowledge and Reading Proficiency for Multilingual Students Across Languages of Instruction. (2025). https://doi.org/10.1037/dev0001858
23. M. Alhamami, Instructional Communication and Medium of Instruction: Content Instructors' Perspectives. (2023). https://doi.org/10.1177/21582440231172713
24. P. N. Iwuanyanwu, Addressing Common Deficiencies of Mathematics Skills Among Chemistry Student Teachers. (2021). https://doi.org/10.4314/ajesms.v17i1.1.
25. J. Hiebert and D. A. Grouws, The Effects of Classroom Mathematics Teaching on Students' Learning. Second Handbook of Research on Mathematics Teaching and Learning. **2**, 371–404 (2007).
26. J. Killpatrick, J. Swafford, and B. Findell, Adding It Up: Helping Children Learn Mathematics. Washington DC: National Academy Press. (2001)
27. M. Kunter, U. Klusmann, J. Baumert, D. Richter, T. Voss, and A. Hachfeld, Professional competence of teachers: Effects on instructional quality and student development. J Educ Psychol. **105**, 805–820 (2013). https://doi.org/10.1037/a0032583
28. D. Loewenberg Ball, M. H. Thames, and G. Phelps, Content knowledge for teaching: What makes it special?. J Teach Educ. **59**, 389–407 (2008). https://doi.org/10.1177/0022487108324554
29. J. Baumert et al, Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. Am Educ Res J. **47**, 133–180 (2010). https://doi.org/10.3102/0002831209345157
30. S. Kholifah, S. B. Waluya, and I. Kharisudin, Student's Mathematics Communication Skill by Reading Comprehension Skill on Solving Word Problem. AIP Conf Proc. **080005**, (2024). https://doi.org/10.1063/5.0133790
31. R. M. Zbiek, M. K. Heid, G. W. Blume, and T. P. Dick, Research on technology in mathematics education: A perspective of constructs. Second Handbook of Research on Mathematics Teaching and Learning. **2**, 1169–1207 (2007)
32. F. Ligozat and Y. Buyck, Comparative Didactics: Toward a Generic Model for Analyzing Content-Specific Dimensions of Teaching Quality. (2024). https://doi.org/10.1177/14749041241257282
33. G. Brousseau, Theory of Didactical Situations in Mathematics. Springer Dordrecht. (2002). https://doi.org/10.1007/0-306-47211-2
34. Y. Chevallard and M. Bosch, Didactic Transposition in Mathematics Education. (1989). https://doi.org/10.1007/978-3-030-15789-0_48.
35. M. Artigue, Learning mathematics in a CAS environment: The genesis of a reflection about instrumentation and the dialectics between technical and conceptual work. International Journal of Computers for Mathematical Learning. **7**, 245–274 (2002). https://doi.org/10.1023/A:1022103903080.
36. J.-B. Lagrange, M. Artigue, C. Laborde, and L. Trouche, Technology and Mathematics Education: A Multidimensional Study of the Evolution of Research and Innovation. Technologuyand Mathematics Education. (2003). https://doi.org/10.1007/978-94-010-0273-8_9.
37. L. Benton, C. Hoyles, I. Kalas, and R. Noss, Bridging Primary Programming and Mathematics: Some Findings of Design Research in England. Digital Experiences in Mathematics Education. **3**, 115–138 (2017). https://doi.org/10.1007/s40751-017-0028-x.
38. N. Calder, Using Scratch: An integrated problem-solving approach to mathematical thinking. Australian Primary Mathematics Classroom. **15**, 9–14 (2010).
39. L. C. Zhang and J. Nouri, A systematic review of learning computational thinking through Scratch in K-9. Comput Educ. **141**, (2019). https://doi.org/10.1016/j.compedu.2019.103607

40. D. Pratt, Making sense of the total of two dice. J Res Math Educ. **31**, 602–625 (2000). https://doi.org/10.2307/749889.
41. C. Konold et al, TinkerPlots Workshop Guide. (2007).
42. A. Yadav, H. Hong, and C. Stephenson, Computational Thinking for All: Pedagogical Approaches to Embedding 21st Century Problem Solving in K-12 Classrooms. TechTrends. **60**, 565–568 (2016). https://doi.org/10.1007/s11528-016-0087-7.
43. UNESCO, Education For Sustainable Development Goals. France. (2017)
44. S. Grover and R. Pea, Computational Thinking in K-12: A Review of the State of the Field. Educational Researcher. **42**, 38–43 (2013). https://doi.org/10.3102/0013189X12463051
45. P21, Partnership for 21st Century Learning A Network of Battelle For Kids Frameworks for 21st Century Learning Definitions. (2019).
46. N. A. H. Hanafiah, Integrating Service Learning in Higher Education for Enhanced Environmental Consciousness in Yan, Kedah. (2024). https://doi.org/10.32890/jgd2024.20.2.2
47. T. Nurul Izza, Sugiman, and S. Kholifah, Digital Didactical Design: Scratch Assisted Learning of Probability Materials in Junior High School. International Journal of Multicultural and Multireligious Understanding. **11**, 277–290 (2024). https://doi.org/10.18415/ijmmu.v11i12.6505
48. M. Akram, Z. Kamran, and M. I. Malik, Maximizing Subject Matter Knowledge of Science Teachers Through Professional Development Plan and Training. (2024). https://doi.org/10.35484/ahss.2024(5-i)07
49. J. W. Creswell and J. D. Creswell, Research Design: Qualitative, Quantitative, and Mixed Methods Approaches. (2018). https://doi.org/10.4324/9780429469237
50. M. Laurent et al, Impact of a programming activity impact of programming on primary mathematics learning. Learn Instr. **82**, (2022). https://doi.org/10.1016/j.learninstruc.2022.101667
51. R. García-Perales and A. Palomares-Ruiz, Education in programming and mathematical learning: Functionality of a programming language in educational processes. Sustainability (Switzerland). **12**, (2020). https://doi.org/10.3390/su122310129
52. L. Benton, P. Saunders, I. Kalas, C. Hoyles, and R. Noss, Designing for learning mathematics through programming: A case study of pupils engaging with place value. Int J Child Comput Interact. **16**, 68–76 (2018). https://doi.org/10.1016/j.ijcci.2017.12.004
53. N. C. Obiuto, W. Ebirim, N. Ninduwezuor-Ehiobu, E. C. Ani, K. A. Olu-lawal, and E. D. Ugwuanyi, Integrating Sustainability Into HVAC Project Management: Challenges and Opportunities. (2024). https://doi.org/10.51594/estj.v5i3.943
54. H. Wang and M. Liu, Methods and Content Innovation Strategies of Digital Education in Higher Vocational Colleges Under the Background of Artificial Intelligence. (2025). https://doi.org/10.1177/14727978251321337
55. M. U. Shahzad, Core Competencies for Digital Leadership Development: A Perspective From the Lens of Paradox Theory. (2024). https://doi.org/10.1108/bl-10-2023-0278
56. A.-R. Smolander and S. R. Yusof, Innovative ICT Education in Early Childhood: Building Young Talents for the Digital World. (2023). https://doi.org/10.37698/jels.v2i1.194
57. S. Siswadi, T. Tiwan, and A. Dharin, Project Based Learning (PjBL) for Improving Elementary School Mathematics Learning Outcomes in Banyumas Regency, Indonesia. (2024). https://doi.org/10.53555/kuey.v30i4.691
58. K. C. Li and B. T. M. Wong, The Opportunities and Challenges of Social Media in Higher Education: A Literature Review. (2021). https://doi.org/10.1007/s42979-021-00857-5
59. B. A. Serrano et al, The Role of FAIR Nanosafety Data and Nanoinformatics in Achieving the UN Sustainable Development Goals: The NanoCommons Experience. (2024). https://doi.org/10.1039/d3su00148b
60. C. N. Ugwu et al, Sustainable Development Goals (SDGs) and Resilient Healthcare Systems: Addressing Medicine and Public Health Challenges in Conflict Zones. (2025). https://doi.org/10.1097/md.0000000000041535