



Students' Achievement in Similar Triangles Using Game-Based Formative Assessment: A Quasi-Experimental Study

Francis Joko E. Guira¹, Alaisah A. Cayana², Crizalyn F. Silos³, Daisy O. Muyano⁴

^{1,2,3,4}Republic Central Colleges, Angeles City, Philippines

Corresponding email: fguira@spcf.edu.ph

ABSTRACT

This study compared Grade 9 students' achievement in geometric proportion and triangle similarity following formative assessment through Quizizz, Gimkit, or Google Forms. A pre-test–post-test comparison involved two intact classes at a private school in Angeles City, Philippines, during School Year 2021–2022. Classes were randomly assigned to either the experimental or control group. A total of 58 students completed both assessments, with 27 in the experimental group and 31 in the control group. The intervention lasted two weeks and included four formative assessment sessions. Both groups received equivalent lesson content, teacher instruction, and assessment questions and followed the same schedules. Achievement was measured using researcher-developed 50-item multiple-choice assessments aligned with a table of specifications and reviewed by mathematics teachers. Independent- and paired-samples t-tests showed similar pre-test scores and significant improvements in both groups, with medium effect sizes. The experimental group had a higher post-test average and greater improvement, but the post-test difference between groups was not significant and had a small effect size. These findings indicate improvement under both assessment conditions without establishing an additional achievement benefit from game-based platforms. The short intervention and inclusion of only one class per condition limit interpretation and generalizability. Mathematics teachers should select assessment platforms according to instructional objectives, learner access, and classroom conditions, while prioritizing aligned questions, informative feedback, and teacher facilitation to support students' mathematical learning.

ARTICLE INFO

Received: Aug. 04, 2026

Revised: Sept. 16, 2026

Accepted: Sept. 21, 2026

KEYWORDS

digital formative assessment, game-based assessment, mathematics achievement, secondary mathematics education, triangle similarity, quasi-experimental design

Suggested Citation (APA Style 7th Edition):

Guira, F.J.K., Cayanan, A.A., & Muyano, D.O. (2026). Students' Achievement in Similar Triangles Using Game-Based Formative Assessment: A Quasi-Experimental Study. *International Research Journal of Science, Technology, Education, and Management*. 6(3), 37-44. <http://doi.org/10.5281/zenodo.23033526>.

INTRODUCTION

Similar triangles are an important component of secondary geometry because the topic requires students to recognize proportional relationships, identify conditions of similarity, and apply geometric principles to measurement problems. These tasks demand both conceptual understanding and procedural accuracy. When learners have difficulty coordinating ratios, diagrams, and theorem-based reasoning, their knowledge may remain fragmented, and their achievement may be limited.

Formative assessment can address these difficulties by giving learners repeated opportunities to answer questions, identify misconceptions, receive feedback, and revise solution strategies while instruction is ongoing. In mathematics, this process is especially valuable because students often need guided practice before they can independently apply abstract concepts. Digital platforms have expanded the ways teachers administer such assessments, particularly in online and flexible learning contexts where immediate monitoring and feedback are necessary (Cahapay & Labrador, 2022; Cassibba et al., 2021).

Digital assessment tools vary in how they structure learning activities. Google Forms allows teachers to present questions, gather responses, and review scores. Quizizz, now called Wayground, is an online learning and assessment platform that offers interactive quizzes with game features such as points, rankings, and immediate feedback. In a Quizizz game-based activity, students answer lesson-related questions through this interactive format (Wayground, n.d.). Gimkit is an online game-based learning platform that combines academic questions with gameplay. Students earn virtual currency or resources by answering correctly and use these rewards to progress in their chosen game mode (Gimkit, n.d.). Both platforms integrate assessment questions with game features that may promote participation and attention. However, their educational value depends on how well the activities align with the lessons, how they are designed, and the classroom conditions in which they are used.

Research generally supports the instructional potential of gamification and digital game-based learning. Gamification has demonstrated positive effects on cognitive, motivational, and behavioral outcomes, although effect sizes vary according to game design and implementation (Sailer & Homner, 2020). Meta-analytic evidence also indicates that digital game-based approaches can improve achievement in science, technology, engineering, and mathematics learning (Wang et al., 2022). Studies involving Quizizz and related platforms further associate game-based activities with engagement, positive learning experiences, and improved mathematical performance (Goksun & Gursoy, 2019; Setiyani et al., 2020; Zhao, 2019). An engaging platform is not necessarily more effective than a familiar digital assessment tool. Research comparing Kahoot, Quizizz, and Google Forms shows differences in students' perceptions of concentration, enjoyment, motivation, and learning experience (Chaiyo & Nokham, 2017). Although game features may capture attention, time limits, competition, unstable internet connections, limited device access, and unfamiliar interfaces can create challenges. Platform effectiveness should therefore be assessed based on the quality of assessment and feedback, ease of access, and teacher guidance, rather than simply how new or exciting it appears.

The study was informed by formative assessment and digital game-based learning perspectives. From a formative assessment perspective, achievement may improve when assessment activities generate evidence of learning and provide opportunities for practice and correction. Digital game-based learning further proposes that interaction, challenge, feedback, and visible progress can support attention and cognitive engagement when meaningfully integrated into instruction (Vlachopoulos & Makri, 2017; Wang et al., 2022). Both conditions in this study provided formative practice, but the experimental condition added game-based features through Quizizz and Gimkit.

Despite the growing use of educational applications, classroom-based evidence directly comparing game-based formative assessment with Google Forms remains limited for specific secondary mathematics topics. Similar triangles provide an appropriate context for comparison because successful performance requires learners to integrate geometric and proportional reasoning. Examining the tools under equivalent lesson content and assessment tasks can clarify whether game-based features produce achievement gains beyond those associated with structured digital

formative assessment itself. Accordingly, this study examined the effects of game-based educational applications as formative assessment tools on Grade 9 students' achievement in similar triangles. It compared students who completed formative assessments through Quizizz and Gimkit with those who used Google Forms within the same online and flexible mathematics learning context. The findings provide practical evidence for teachers and school leaders selecting digital assessment tools to support measurable learning outcomes.

RESEARCH OBJECTIVES

This study examined Grade 9 students' achievement in similar triangles when game-based educational applications were used for formative assessment. Specifically, it aimed to describe the control and experimental groups' pre-test scores and determine whether these differed significantly, describe both groups' post-test scores, assess whether each group improved significantly from pre-test to post-test, and compare the post-test scores of students who used game-based applications with those who used Google Forms.

MATERIALS AND METHODS

Research Design and Locale

The study used a quasi-experimental pre-test-post-test design with one control group and one experimental group. The design was appropriate because existing Grade 9 sections were retained and individual random assignment was impractical in the school setting. The study was conducted at Republic Central Colleges in Angeles City, Philippines, during the third quarter of School Year 2021-2022.

Participants and Group Assignment

The study followed a quasi-experimental pre-test-post-test design with one control group and one experimental group. This design was suitable because existing Grade 9 classes were kept intact, as randomly assigning individual students was impractical in the school setting. The study took place at Republic Central Colleges in Angeles City, Philippines, during the third quarter of School Year 2021–2022.

Research Instrument

Students' achievement was measured using a researcher-developed 50-item multiple-choice test comprising 17 items on geometric proportion and 33 items on triangle similarity. A two-way table of specifications guided the alignment of content coverage, learning competencies, and intended cognitive levels (Fives & DiDonato-Barnes, 2013). The pre-test and post-test contained different questions covering the same topics to reduce students' recall of earlier questions. Three mathematics teachers evaluated the instrument for clarity, grammar, alignment with the content, and suitability. Their feedback guided revisions before pilot testing. The pre-test was subsequently pilot-tested among 39 Grade 9 students from another section and yielded a Kuder–Richardson Formula 20 coefficient of .69, interpreted cautiously as marginal internal consistency because it fell slightly below the commonly cited .70 benchmark (Tavakol & Dennick, 2011). Item difficulty analysis identified 35 average items that were retained and 15 items flagged for rejection, comprising 11 very difficult and four easy items. The flagged items were subsequently revised or replaced, maintaining the final test at 50 items and preserving the allocation across the two content domains. The reported reliability coefficient pertains to the pilot-tested version before these subsequent changes.

Intervention and Data Collection Procedure

The intervention lasted two weeks and comprised four formative assessment sessions, with 10 items administered per session. The first two sessions addressed geometric proportion, while the remaining sessions focused on triangle similarity. The same mathematics teacher delivered the lessons to both groups, while the researchers prepared the assessment materials and administered the activities under the supervision of the teacher in charge. Both groups received the same lesson content, assessment questions, and instructional schedule. Sessions were conducted synchronously through Google Meet during the students' scheduled quiz periods. The experimental group completed game-based activities through Quizizz or Gimkit, whereas the control group answered the same questions through Google Forms. Attendance was checked before each assessment to monitor participation. When students encountered connectivity problems or difficulties navigating the platforms, the teacher and researchers provided technical assistance. Scores were released upon completion of the assessments. The 50-item pre-test and post-test were administered through Google Forms, with one hour allotted for each test under comparable administration conditions.

Data Analysis

Frequency and percentage were used to summarize the distribution of achievement scores, while mean and standard deviation described students' performance on the 50-item test. Percentage correct was calculated by dividing each student's raw score by 50 and multiplying by 100, with higher scores indicating greater achievement. Although the Department of Education (2012) ranks Advanced above Proficient, its grading framework does not establish the original 10-point raw-score intervals used in this study; therefore, these unsupported proficiency classifications were removed. Independent-samples t-tests compared the groups' pre-test and post-test scores, while paired-samples t-tests assessed changes within each group. Cohen's *d* was reported for independent comparisons, and Cohen's *d_z* was reported for paired comparisons. All statistical tests were evaluated at the .05 significance level. Normality and homogeneity of variance were assessed using the Shapiro–Wilk and Levene tests, respectively.

Ethical Considerations

Approval from school authorities was obtained before data collection. Students and their parents or legal guardians were informed about the study's purpose, procedures, confidentiality measures, and voluntary participation, including the right to withdraw without academic penalty. Parents or legal guardians provided written informed consent, while students gave their assent before participating. These conditions were recorded in the consent and assent forms. Findings were presented as group results, and personal information was managed in accordance with the Data Privacy Act of 2012 (Republic Act No. 10173, 2012).

RESULTS AND DISCUSSION

Pre-Test Performance Prior to Implementation

Before the intervention, most students in both groups scored between 11 and 20 on the 50-item achievement test. The control group had an average score of 17.77 (*SD* = 3.88), while the experimental group averaged 17.30 (*SD* = 4.91). No student scored above 30, indicating relatively low initial achievement in geometric proportion and triangle similarity. These score ranges summarize the distribution of scores and do not represent validated proficiency levels. An independent-samples t-test showed that the groups' pre-test scores did not differ significantly, $t(56) = 0.414$, $p = .681$. The effect size was small (Cohen's *d* = 0.11). Although the groups had similar mean pre-test scores, the nonsignificant result does not establish statistical equivalence. As Amrhein et al. (2019) emphasized, a statistically nonsignificant result should not be interpreted as proof that no difference exists. The baseline findings therefore provide context for interpreting subsequent achievement scores without ruling out other differences between the groups. Furthermore, having only one intact class per condition limits the separation of intervention effects from class-specific influences (see Table 1).

Table 1. Pre-Test Distribution of Scores for the Control and Experimental Groups

Score Category	Control, n (%)	Experimental, n (%)
Advanced (41-50)	0 (0.0)	0 (0.0)
Proficient (31-40)	0 (0.0)	0 (0.0)
Approaching Proficiency (21-30)	7 (22.6)	7 (25.9)
Developing (11-20)	23 (74.2)	16 (59.3)
Beginning (0-10)	1 (3.2)	4 (14.8)
Mean	17.77	17.30
Standard deviation	3.88	4.91
Verbal interpretation	Developing	Developing

Note. Achievement categories were based on the 50-item test: Advanced (41-50), Proficient (31-40), Approaching Proficiency (21-30), Developing (11-20), and Beginning (0-10).

Both groups had higher average achievement scores after the intervention. The control group had a post-test mean of 21.45 (SD = 7.33), with 45.2% scoring between 21–30 and 12.9% between 31–40. The experimental group averaged 23.41 (SD = 7.88), with 48.1% scoring between 21–30, 7.4% between 31–40, and 7.4% between 41–50. Although only the experimental group had students in the highest score range, these descriptive results do not demonstrate that its approach was more effective. The ranges summarize score distributions and do not represent validated proficiency levels (see Table 2).

The improvements may be related to learning activities shared by both groups, such as teacher discussions, practice, and feedback. Wisniewski et al. (2020) found that feedback helps learning, especially when it gives students useful guidance for improvement. Sailer and Homner (2020) and Wang et al. (2022) also reported positive average effects of gamification and digital game-based instruction on learning outcomes. Although these studies suggest potential benefits of Quizizz and Gimkit, the present findings cannot separate the contribution of game features from other instructional factors. Engagement and familiarity with the platforms were not directly measured, so explanations involving these factors remain uncertain. These descriptive findings should therefore be interpreted alongside statistical comparisons within and between the groups.

Table 2. Post-Test Distribution of Scores for the Control and Experimental Groups

Score Category	Control, n (%)	Experimental, n (%)
Advanced (41-50)	0 (0.0)	2 (7.4)
Proficient (31-40)	4 (12.9)	2 (7.4)
Approaching Proficiency (21-30)	14 (45.2)	13 (48.1)
Developing (11-20)	12 (38.7)	9 (33.3)
Beginning (0-10)	1 (3.2)	1 (3.7)
Mean	21.45	23.41
Standard deviation	7.33	7.88
Verbal interpretation	Approaching Proficiency	Approaching Proficiency

Note. The same achievement categories were applied to the post-test scores.

Within-Group Improvement in Achievement

Both groups showed statistically significant gains in achievement. The control group's average score increased by 3.68 points, $t(30) = 3.514$, $p = .001$, Cohen's $d_z = 0.63$, while the experimental group's average increased by 6.11 points, $t(26) = 3.716$, $p = .001$, Cohen's $d_z = 0.72$. Both gains reflected medium within-group effects. Although the experimental group improved more descriptively, the paired-samples tests do not show whether its gain was significantly greater than the control groups. The improvements also cannot be attributed only to the platforms, since both groups received teacher-led instruction, practice, and feedback (see Table 3).

These findings align with research on feedback and game-based learning. Wisniewski et al. (2020) found that feedback supports learning, especially when it clarifies errors and guides improvement. Sailer and Homner (2020) and Wang et al. (2022) also reported positive average effects of gamification and digital game-based instruction. Shared instructional activities may have supported improvements in both groups, while game features may have encouraged participation. However, engagement was not measured directly, so this explanation remains tentative. Game-based activities may therefore complement questions aligned with lesson goals, useful feedback, and teacher guidance.

Table 3. Paired-Samples t-Tests Comparing Pre-Test and Post-Test Scores Within Groups

Group	N	Pre-Test M	Post-Test M	Mean Gain	$t(df)$	p	Cohen's d_z	Verbal Interpretation
Control	31	17.77	21.45	3.68	3.514 (30)	.001	0.63	Medium effect
Experimental	27	17.30	23.41	6.11	3.716 (26)	.001	0.72	Medium effect

Note. Mean gain was computed as post-test minus pre-test. Cohen's d_z was derived from the paired-samples t statistic and sample size.

Between-Group Difference in Post-Test Achievement

The experimental group had a higher average post-test score ($M = 23.41$, $SD = 7.88$) than the control group ($M = 21.45$, $SD = 7.33$), but the difference was not statistically significant, $t(56) = -0.979$, $p = .332$. The small effect size (Cohen's $d = 0.26$) showed a modest observed advantage, but there was insufficient evidence that game-based assessment led to better achievement during the two-week intervention. This finding does not mean that the two approaches were equivalent. While Sailer and Homner (2020) reported positive average effects of gamification, Göksün and Gürsoy (2019) found that outcomes differed across gamified platforms, suggesting that benefits depend on how these tools are used (see Table 4).

Both groups' improvements may reflect shared instructional features, including teacher discussion, aligned questions, practice, and feedback. Wisniewski et al. (2020) emphasized that feedback supports learning particularly through the information it provides. Meanwhile, Wang and Tahir (2020) identified technical difficulties and time pressure as potential challenges in game-based quizzes. Connectivity and navigation difficulties were observed in this study, but their effects on achievement were not measured. Teachers should therefore select platforms according to instructional objectives, student access, and classroom conditions while maintaining clear instructions and informative feedback. Combining game-based activities with Google Forms may be considered, although the effectiveness of a combined approach was not tested.

Table 4. Independent-Samples t-Test Comparing Post-Test Scores of the Control and Experimental Groups

Group	N	Mean	SD	<i>t</i> (56)	<i>p</i>	Cohen's <i>d</i>	Verbal Interpretation
Control	31	21.45	7.33	-0.979	.332	0.26	Small Effect
Experimental	27	23.41	7.88	-	-	-	

*Note. Equal variances were assumed based on Levene's test, $F = 0.013$, $p = .909$. Cohen's *d* is reported as a positive magnitude favoring the experimental group.*

CONCLUSION AND RECOMMENDATION

This study examined Grade 9 students' achievement in similar triangles when game-based applications were used for formative assessment. The control and experimental groups had similar initial scores. After the two-week intervention, both groups showed statistically significant improvements with medium within-group effect sizes. Although the experimental group had a higher average post-test score and a larger average gain, the post-test difference between groups was not statistically significant and had a small effect size.

The findings suggest that Quizizz and Gimkit can function as effective formative assessment tools, but they were not statistically superior to Google Forms under the conditions examined. The result highlights the importance of instructional alignment, question quality, feedback, accessibility, student familiarity, and teacher facilitation. Game elements may support attention and participation, but learning outcomes depend on how the platform is integrated into the teaching and assessment process.

Mathematics teachers are encouraged to use game-based applications purposefully for review, confidence-building, immediate feedback, and checks for understanding. Google Forms may continue to be used for structured, accessible, and lower-pressure practice. Combining the two approaches may enable teachers to balance engagement with accessibility and select the format that best matches the instructional objective, task difficulty, and learner readiness. School leaders may also provide professional development that emphasizes assessment design and interpretation rather than platform operation alone.

Study Limitation

This study was limited to 58 students from two intact Grade 9 classes in one private school, restricting generalizability and the detection of small between-group differences. Having only one class per condition prevented the separation of intervention effects from class-specific influences, while student-level analyses did not account for classroom clustering. Excluding five students with incomplete assessments may also have introduced bias. The two-week intervention measured immediate achievement without assessing long-term retention. Measurement limitations included marginal pilot-test reliability ($KR-20 = .69$). Other limitations included the lack of reported reliability estimates after the items were revised and the lack of evidence that the pre-test and post-test were equally difficult. Internet connection problems and difficulties navigating the platforms were observed, but their impact was not measured. Engagement, motivation, and familiarity with the platforms were also not assessed. Using Quizizz and Gimkit within one condition also prevented identification of their individual contributions. Consequently, the findings do not establish platform-specific causality or equivalence between approaches. Future research should involve multiple classes across schools, longer interventions, comparable and reliable assessment forms, analyses accounting for classroom clustering and baseline achievement, and delayed retention measures.

REFERENCES

- Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, 567(7748), 305–307. <https://doi.org/10.1038/d41586-019-00857-9>
- Cahapay, M. B., & Labrador, M. G. P. (2022). Instructional practices of Filipino teachers in remote mathematics education during the COVID-19 pandemic. *International Journal of Didactical Studies*, 3(1), Article 101458. <https://doi.org/10.3390/IJODS.202211753>
- Cassibba, R., Ferrarello, D., Mammanna, M. F., Musso, P., Pennisi, M., & Taranto, E. (2021). Teaching mathematics at distance: A challenge for universities. *Education Sciences*, 11(1), Article 1. <https://doi.org/10.3390/educsci11010001>
- Chaiyo, Y., & Nokham, R. (2017). The effect of Kahoot, Quizizz and Google Forms on the student's perception in the classrooms response system. In *2017 International Conference on Digital Arts, Media and Technology (ICDAMT)* (pp. 178–182). IEEE. <https://doi.org/10.1109/ICDAMT.2017.7904957>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Data Privacy Act of 2012, Republic Act No. 10173 (2012) (Phil.). <https://privacy.gov.ph/data-privacy-act/>
- Department of Education. (2012). *Guidelines on the assessment and rating of learning outcomes under the K to 12 basic education curriculum* (DepEd Order No. 73, s. 2012). https://www.depednegor.net/uploads/8/3/5/2/8352879/rating_of_learning_outcomes.pdf
- Fives, H., & DiDonato-Barnes, N. (2013). Classroom test construction: The power of a table of specifications. *Practical Assessment, Research, and Evaluation*, 18, Article 3. <https://doi.org/10.7275/cztt-7109>
- Gimkit. (n.d.). *Gimkit*. <https://www.gimkit.com/>
- Göksün, D. O., & Gürsoy, G. (2019). Comparing success and engagement in gamified learning experiences via Kahoot and Quizizz. *Computers & Education*, 135, 15–29. <https://doi.org/10.1016/j.compedu.2019.02.015>
- Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, 32, 77–112. <https://doi.org/10.1007/s10648-019-09498-w>
- Setiyani, S., Fitriyani, N., & Sagita, L. (2020). Improving students' mathematical problem-solving skills through Quizizz. *JRAMathEdu (Journal of Research and Advances in Mathematics Education)*, 5(3), 276–288. <https://doi.org/10.23917/jramathedu.v5i3.10696>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Vlachopoulos, D., & Makri, A. (2017). The effect of games and simulations on higher education: A systematic literature review. *International Journal of Educational Technology in Higher Education*, 14, Article 22. <https://doi.org/10.1186/s41239-017-0062-1>
- Wang, A. I., & Tahir, R. (2020). The effect of using Kahoot! for learning—A literature review. *Computers & Education*, 149, Article 103818. <https://doi.org/10.1016/j.compedu.2020.103818>
- Wang, L. H., Chen, B., Hwang, G. J., Guan, J. Q., & Wang, Y. Q. (2022). Effects of digital game-based STEM education on students' learning achievement: A meta-analysis. *International Journal of STEM Education*, 9, Article 26. <https://doi.org/10.1186/s40594-022-00344-0>
- Wayground. (n.d.). *Free online quiz maker*. <https://wayground.com/home/quiz-maker>
- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, Article 3087. <https://doi.org/10.3389/fpsyg.2019.03087>
- Zhao, F. (2019). Using Quizizz to integrate fun multiplayer activity in the accounting classroom. *International Journal of Higher Education*, 8(1), 37–43. <https://doi.org/10.5430/ijhe.v8n1p37>