Journal of Practical Studies in Education
ISSN: 2634-4629

Enhancing Critical Thinking Skills through Computer-Aided Assessment on Non-Routine Problem Solving across Three Distinct Learning Delivery Modalities

Marivic G. Molina (Corresponding Author)
Graduate School, Centro Escolar University, Manila, Philippines
College of Science, Pamantasan ng Lungsod ng Maynila, Manila, Philippines
E-mail: mgmolina@plm.edu.ph

Erna V. Yabut
Graduate School, Centro Escolar University, Manila, Philippines

Received: 24/12/2024
Accepted: 26/02/2025
Published: 01/03/2025

Volume: 6 Issue: 2

How to cite this paper: Molina, M. G., & Yabut, E. V. (2025). Enhancing Critical Thinking Skills Through Computer-Aided Assessment on Non-Routine Problem Solving across Three Distinct Learning Delivery Modalities. *Journal of Practical Studies in Education*, *6(*2), 23-32
DOI: https://doi.org/10.46809/jpse.v6i2.98

Abstract

This study investigated the impact of computer-aided assessment on the enhancement of critical thinking skills in non-routine problem-solving across three distinct learning delivery modalities: fully online, balanced blended, and predominantly face-to-face (F2F) blended. The study aimed to evaluate the effectiveness of computer-aided assessment in enhancing students' critical thinking skills through non-routine problem-solving across three distinct learning delivery modalities. A quasi-experimental non-equivalent pre-test and post-test design was employed. Participants were drawn from three higher education institutions, each offering one of the three learning delivery modalities. In each institution, two intact groups were conveniently chosen and then randomly assigned to an experimental group that engaged with computer-aided assessment and a control group that used printed or digital problem sets. Data were analyzed using mean and standard deviation, paired and independent t-test and two-way ANOVA. Results showed that computer-aided assessment enhanced students' critical thinking skills through non-routine problem-solving, with varying effects across different learning delivery modalities. The balanced blended modality benefited the most from the computer-aided assessment intervention and proved most effective, showing the highest and most consistent gains, while the face-to-face blended and fully online modalities showed different patterns of improvement. These results suggest the importance of tailoring instructional strategies to specific learning environments to maximize the impact of computer-aided assessment.

Keywords: Critical Thinking Skills, Computer-Aided Assessment, Non-Routine Problems, Learning Delivery Modalities, Blended Learning, Online Learning

1. Introduction

In an ever-evolving educational landscape, critical thinking has emerged as one of the most essential skills of the 21st century (Dwi Susandi et al., 2019; Hafni et al., 2019; Hujjatusnaini et al., 2022; Mahanal et al., 2019; Yumiati & Kusumag, 2019). Paul (1993, as cited in Wang & Abdullah, 2024) defined critical thinking as a disciplined process that involves the active and skillful conceptualization, analysis, synthesis, and evaluation of information derived from observation, experience, reasoning, or communication, all of which guide belief and action. Similarly, Aini et al. (2019) argued that critical thinking distinguishes students who simply perform mathematical procedures from those who genuinely understand the reasoning behind their actions. Critical thinking allows students to choose the most effective strategies from a range of potential approaches to solve problems and navigate the complex challenges encountered in life (Demirel et al., 2017; Zukhairina et al., 2020).

According to Scriven and Paul (2005), critical thinking is not an innate ability and does not develop automatically. Instead, it is a skill that must be learned and refined through consistent practice. Students with strong critical thinking skills approach problem-solving with greater care, leading to more accurate conclusions and logical solutions (Berestova et al., 2022). Similarly, Angeli and Valanides (2009, as cited in Rahmasari et al., 2023) emphasized that individuals with well-developed critical thinking skills are better equipped to handle large volumes of information and solve problems that lack clear-cut solutions. Such problems, often categorized as non-routine, require more than basic procedural knowledge.

Non-routine problems (NRPs) are those that cannot be solved through the simple application of established rules or procedures (Laset & Limjap, 2005, as cited in Andrade et al., 2020). Many students struggle with NRPs due to limited exposure to this type of problem (Yeo, 2009, as cited in Gavaz et al., 2021). This struggle is partly because only a small portion of problems in textbooks are non-routine (Berisha, 2015; Fan & Zhu, 2000; Kablan & Uğur, 2019; Kolovou et al., 2009; Manopo & Lisarani, 2021; van Zanten & van den Heuvel-Panhuizen, 2018). Introducing NRPs to all students is crucial, as it helps meet the demands of the modern workplace, where critical and creative thinking are increasingly valued (OECD, 2010, as cited in Andrade et al., 2020). However, preparing students for NRPs has become more challenging, particularly due to the disruptions caused by the COVID-19 pandemic (Julie et al., 2017; Tanu Wijaya, 2020).

One significant challenge in exposing students to NRPs is the need for timely feedback, which is difficult in large classrooms of 30 to 40 students. Providing individual feedback requires substantial effort from teachers, and many instructors only provide feedback at the end of the term (Ajogbeje, 2023). To address this issue, the integration of technology, such as computer-aided assessment (CAA), has become essential. CAA systems, with features like algorithms, grading codes, adaptivity, and feedback capabilities, are particularly useful for formative assessment and can efficiently provide individualized feedback in larger classrooms (Barana, Conte, et al., 2018; Barana, Marchisio, et al., 2021; Kundu & Bej, 2020). Abulhul (2021) noted that integrating technology into the classroom not only improves the learning environment but also promotes students' critical thinking and enhances their ability to apply knowledge effectively.

The rise of digital tools, accelerated by the COVID-19 pandemic, has led to the increased popularity of online and blended learning. Online learning allows interactions between students and instructors through online platforms, providing flexible and accessible learning experiences (Natarajan et al., 2022; Rural et al.,2022). Blended learning, which combines online and offline methods, offers a comprehensive learning experience (Suana et al., 2019). According to Tong et al. (2023), blended learning combines traditional in-person teaching with technology-enhanced instruction to meet changing educational demands.

Given these developments, it is essential to foster critical thinking skills in students to ensure their success in tackling non-routine problems through a technology-enhanced learning environment, which can be delivered either through blended or online methods. In light of this, the primary objective of this study was to evaluate the effectiveness of computer-aided assessment (CAA) in enhancing students' critical thinking skills through non-routine problem-solving across three distinct learning delivery modalities: balanced blended, predominantly F2F blended and fully online.

1.1. Research Objectives

The objectives of this study were:

1. To evaluate the effectiveness of computer-aided assessment (CAA) in enhancing students' critical thinking skills through non-routine problem solving.
2. To examine whether the impact of CAA on critical thinking skills through non-routine problem solving varies across balanced blended, predominantly face-to-face (F2F) blended, and fully online learning delivery modalities.
3. To identify which learning delivery modality benefited the most from the CAA in enhancing students' critical thinking skills through non-routine problem solving.

1.2. Significance of the Study

This study was important as it explored the impact of computer-aided assessment (CAA) in enhancing critical thinking skills, a vital competency in the 21st century. As technology becomes increasingly integrated into education, understanding the influence of CAA on students' ability to engage in non-routine problem-solving is essential for designing effective instructional strategies. Furthermore, by analyzing its effects across different learning delivery modalities, this research provided valuable insights into the most efficient ways for integrating technology-based assessments. The findings support educators, policymakers, and curriculum developers in making informed decisions about instructional design, ultimately enhancing students' critical thinking skills and overall learning experiences in various educational settings.

2. Method

2.1. Research Design

The study employed a quasi-experimental research design. According to Jhangiani et al. (2019), this type of design is similar to an experimental research design but does not fully satisfy the requirements of a true experimental study. Specifically, it lacks one of the essential components of experimental research: random assignment. Although the independent variable is manipulated, either a control group is absent, or participants are not randomly assigned to groups (Cook & Campbell, 1979, as cited in Jhangiani et al., 2019). This design was suitable for the present study because the professors who agreed to involve their students determined which intact classes would participate, resulting in the absence of individual-level random assignment. Even though the two pre-existing groups were subsequently assigned randomly to the experimental and control groups, the lack of randomization at the individual level remained a defining characteristic of the quasi-experimental design. While the random assignment of intact groups helped mitigate potential biases and enhance comparability, some limitations to internal validity persisted.

The study compared an experimental group (EG) against a control group (CG), where the EG used the CAA to solve the NRPs, and the CG received either printed or digital problem sets containing the same NRPs within the CAA. The EG used the CAA to solve the NRPs, while the CG received either printed or digital problem sets containing the same NRPs within the CAA.

Specifically, the study used a Pretest-Posttest Non-equivalent Groups Design with two groups: a treatment group and a control group (Jhagiani et al., 2019). The treatment group took a pretest, received the treatment, and then took a posttest. The control group also took a pretest and posttest but did not receive the treatment. The goal was to see not only if the treatment group improved but whether their improvement was greater than that of the control group.

In this study, both groups took a pretest; however, the experimental group (EG) used the CAA before taking the posttest, while the control group (CG) did not. This design assessed whether participants who utilized the CAA showed improvement and whether their improvement exceeded that of those who had not used it.

2.2. Participants and Sampling

The study involved students from three different types of higher education institutions in Manila, Philippines: a local university, a sectarian college, and a non-sectarian university. Each institution used a distinct learning modality: a local university with balanced blended (50% face-to-face (F2F) and 50% online) modality, a sectarian college with predominantly F2F blended (80% F2F and 20% online) modality, and a non-sectarian university with fully online (100% online) modality.

Two intact classes from each university participated — one assigned to the experimental group (EG) and the other to the control group (CG) — through simple random sampling. Informed consent was obtained from students, and those under 18 or taking the course for the second time were excluded to maintain a representative undergraduate age range (18-22). EG participants needed devices with stable internet access.

A comparability test was conducted to ensure that both groups were similar in their ability to solve non-routine problems (NRPs) before the intervention. The pretest results showed no significant differences between the EG and CG across all learning delivery modalities, confirming that any differences observed in the posttest results were due to the intervention, and not to pre-existing differences between the groups.

After eligibility screening, informed consent collection, and comparability testing, participant numbers were adjusted. In the balanced blended modality, the EG had 41 students, and the CG had 34. In the predominantly F2F blended modality, the EG had 22 students, and the CG had 24. In the fully online modality, the EG had 27 students, and the CG had 30.

Overall, 178 students took part: 75 in the balanced blended modality, 46 in the predominantly F2F blended modality, and 57 in the fully online modality. Across the study, 90 students were in the EG and 88 in the CG.

2.3. Intervention: Computer-Aided Assessment (CAA)

The Computer-Aided Assessment (CAA), named 'MatHOTSanayan,' was developed using the Design and Development Research (DDR) method. As described by Plomp (2007, as cited in Ghazali et al., 2022), DDR is a systematic approach used to create and refine educational interventions—such as programs, instructional strategies, products, and systems—with the dual purpose of solving complex educational challenges and gaining deeper insights into the design and development processes of these interventions. The DDR method was chosen for the development of the CAA in this study due to its structured methodology, which ensures that the CAA's development is both systematic and responsive to real-world educational needs.

The development of the CAA followed the three phases of the DDR approach: needs analysis, design and development, and evaluation—an approach adopted by several researchers (Jaya et al., 2021; Noh & Karim, 2021; Padzil et al., 2021).

In the needs analysis phase, literature on NRP-solving, technology-enhanced assessments, and DDR was reviewed, identifying gaps in assessing non-routine problem-solving performance and limitations of traditional methods, highlighting the need for CAA.

In the design and development phase, the CAA was created to engage students in NRP-solving. Specifically designed for the *Mathematics in the Modern World* (MMW) course, it focused on two main topics: Mathematical Language and Symbols, and Problem Solving and Reasoning.

The CAA used formative assessments with open-ended questions requiring manual input, promoting active problem-solving. Platform selection balanced cost efficiency and reliability, and the name, *'MatHOTSanayan,'* combined 'Math,' 'HOTS,' and '*Sanayan*'—a Filipino word for consistent practice. The CAA consisted of four modules, each containing 30 NRPs, which were reviewed and validated by five mathematics experts teaching MMW. Their input helped improve and refine the modules. After finalization, the NRPs and their corresponding answers were programmed into the system, accommodating multiple correct answers and response formats to capture a variety of student inputs. Additionally, detailed instructions and a module overview were prepared and encoded at the beginning of each module.

Key features of the CAA included:

- Immediate Verification Feedback: Provided instant verification and sample solutions for enhanced learning.
- Focus on HOTS: Encouraged critical and creative problem-solving.
- Triple Attempts Opportunities: Allowed up to three tries per problem, promoting persistence and learning from mistakes.
- Skip and Revisit Options: Enabled students to skip and revisit problems after multiple failed attempts.
- Sequential Module Unlocking: Ensured mastery of one topic before progressing to the next.
- Unlimited Access to Unlocked Modules: Allowed continuous review of completed modules for reinforcement.
- Randomized Problem Order: Prevented memorization and encouraged deeper understanding.
- Automated Grading: Provided instant scoring, saving time for both students and teachers.

In the evaluation phase, the CAA was assessed through quantitative and qualitative methods. Expert validation was used for the quantitative method, while user experience was used for the qualitative method. Expert validators rated its functionality, accessibility, technical performance, mobile design, privacy, social presence, teaching presence, and cognitive presence using the Rubric for e-Learning Tool Evaluation, adapted from Anstey and Watson (2018). High ratings were given in most categories, especially in mobile design and cognitive presence, where experts reached unanimous agreement. Minor concerns were noted in social presence, reflecting the CAA's emphasis on self-regulated learning over collaborative activities.

User feedback from 15 students commended the CAA's intuitive interface, cross-device compatibility, triple-attempt opportunities, skip-and-revisit options, and immediate verification feedback. Students especially valued its support for higher-order thinking skills (HOTS), which foster critical thinking, as well as its role in promoting self-paced learning. Some technical challenges and minor design limitations were identified, offering opportunities for improvement. Despite these, the CAA proved reliable, accessible, and effective in fostering individual growth, demonstrating potential for broader adoption in online education.

The suggestions and comments gathered during the evaluation phase were used to refine and enhance the CAA prior to its implementation, ensuring that it effectively serves its purpose.

2.4. Instruments

The study employed a researcher-developed test and rubrics to evaluate students' critical thinking skills and non-routine problem-solving performance. This test was called the Non-Routine Problem-Solving Test (NRPST) and was accompanied by the Non-Routine Problem-Solving Scoring Rubric (NRPSSR) and the Critical Thinking Skills Leveling Rubric (CRITSLR). The NRPST consisted of three non-routine problems (NRPs) from two selected chapters of MMW: mathematical language and symbols, and problem-solving and reasoning. Each item in the NRPST included accompanying questions to encourage critical thinking. The NRPSSR comprised six criteria: understanding the problem, use of mathematical concepts, strategies, computations and procedures, explanation, and reasoning, while the CRITSLR included four criteria: inquiry, analysis, evaluation, and sound reasoning and conclusion.

Table 1 classifies students' mean scores on individual and overall NRPST items, with performance levels ranging from Novice to Expert and critical thinking skills from Not Critical to Very Critical.

Table 1. Classification of Students' Mean Scores for Individual and Overall NRPST Items

Score Range		NRP-solving	CRITS Level
Individual	Overall	Performance	
4.01 – 5.00	12.01 – 15.00	Expert	Very Critical
3.01 – 4.00	9.01 – 12.00	Practitioner	Critical
2.01 – 3.00	6.01 – 9.00	Apprentice	Sufficiently Critical
1.01 – 2.00	3.01 – 6.00	Advanced Beginner	Less Critical
0.00 – 1.00	0.00 – 3.00	Novice	Not Critical

The NRPST and rubrics were validated by five PhD-holding mathematics experts with extensive teaching experience. Their feedback refined the items and ensured alignment with the study's goals. The revised versions were confirmed valid and endorsed for use.

A pilot test with 30 students not included in the main study followed validation. Their responses were independently scored by the researcher and two other professors using the NRPSSR and CRITSLR. Inter-rater reliability was assessed through Cronbach's alpha coefficients, showing strong reliability across all criteria. Understanding the problem ranged from 0.857 to 1.000, use of mathematical concepts from 0.765 to 0.996, strategies from 0.895 to 0.998, computations and procedures from 0.970 to 0.999, explanation from 0.987 to 1.000, and reasoning from 0.985 to 0.997. These coefficients indicated excellent agreement among raters.

The high Cronbach's alpha coefficients across most criteria highlighted the strong reliability of the NRPSSR. The criteria of understanding the problem, strategies, computations and procedures, explanation, and reasoning showed "almost perfect" reliability, indicating excellent agreement among raters. Although the use of mathematical concepts showed some variability, it still achieved "substantial" reliability, reflecting strong rater agreement.

The CRITSLR also demonstrated high reliability. Interpretation ranged from 0.967 to 0.998, analysis from 0.989 to 0.998, evaluation from 0.982 to 0.996, and sound reasoning and conclusion from 0.966 to 0.991, all showing "almost perfect" consistency.

These results confirmed the sound development and reliability of both rubrics, ensuring consistent scoring across raters. The NRPSSR was used only for the pretest to assess non-routine problem-solving performance of the experimental and control groups, establishing their comparability. The CRITSLR was applied to both pretest and posttest to evaluate the effectiveness of CAA in enhancing students' critical thinking skills through non-routine problem-solving.

2.5. Data Collection

The study consisted of three phases of data collection: Phase 1 – Pretest, Phase 2 – Exposure to NRPs, and Phase 3 – Posttest. Before the data collection, the researcher obtained permission from the appropriate university officials at the three higher education institutions through letters of permission. To carry out the study, the researcher coordinated with the professors of the participants from two of the institutions while including the researcher's own two classes at the third institution. Additionally, the researcher requested full consent from the students using an informed consent form.

In Phase 1, the NRPST was administered to both the experimental group (EG) and the control group (CG) during the first week of the semester. The researcher administered the test, and participants completed it individually. To encourage sincere effort, participants were told that their credit would be based on the quality and thoughtfulness of their responses, with only unanswered questions receiving no credit.

In Phase 2, both the EG and CG were exposed to NRPs but through different methods. The EG used the computer-aided assessment (CAA) immediately after the discussion of mathematical language and symbols. Participants accessed the CAA one or two days before the first week of the topic on problem-solving and reasoning and had flexible time over two weeks to work through the CAA both during and after class. They were allowed to retake modules in the CAA in which they struggled to improve their scores, repeating them until they reached their target scores or until the answering period ended. The data generated by the CAA system included performance metrics, response accuracy, the number of attempts per item and per module, and time spent on each item as well as each module, providing comprehensive insights into participants' engagement and progress.

Meanwhile, the CG worked on printed or digital problem sets containing the same NRPs as those in the CAA. These problem sets were distributed during the weeks covering problem-solving and reasoning, with a new problem set provided at each meeting, totaling four problem sets. CG participants completed and submitted their problem sets during class hours, writing their solutions on paper and handing them in at the end of each class. Corrected problem sets were returned in the next session, and any additional work done outside class was not required to be submitted or reviewed by the researcher.

During this time, the researcher acted as a facilitator, available only for questions about instructions for both the EG and CG or technical issues with the CAA for the EG. Both groups solved the NRPs independently.

After two weeks of working with the NRPs, both groups attended two 'Sharing of Strategies' sessions to discuss the methods they used to solve the problems. These sessions helped participants learn different approaches and discover solutions for challenging problems. Only students presented their solutions, while the researcher facilitated discussions and intervened only to correct misinformation or clarify concepts. Depending on the learning modality, sessions were held in person, online, or in a blended format.

In Phase 3, the researcher administered the NRPST as a posttest to both the EG and CG immediately after the 'Sharing of Strategies' sessions. Participants took the test individually and were once again reminded that their effort and careful approach to problem-solving would be acknowledged, while only unanswered questions would receive no credit. They were also reassured that the posttest results would not impact their MMW grades but were encouraged to do their best.

2.6. Data Analysis

The data analysis for this study employed statistical methods to evaluate the effectiveness of computer-aided assessment (CAA) in enhancing students' critical thinking skills through non-routine problem-solving across different learning delivery modalities. Descriptive statistics, including means and standard deviations, were used to summarize students' critical thinking skill levels across the three different learning delivery modalities and between the experimental group (EG) and the control group (CG). Additionally, inferential statistics, such as a paired t-test, were conducted to assess pre- and post-test critical thinking skills score improvements within the EG, while an independent t-test was used to assess differences in post-test critical thinking skills score between the EG and CG. Two-way ANOVA and post-hoc tests (Tukey HSD) were applied to determine whether there were significant differences in critical thinking skills between the EG and CG and among the three learning delivery modalities.

2.7. Ethical Considerations

The CEU Institutional Ethics Review Board (IERB) approved the study, ensuring strict adherence to ethical considerations. Informed consent was obtained from all participants, providing clear information on the study's purpose, procedures, potential risks, benefits, and participants' rights, including the right to withdraw at any time without affecting their MMW grades. Privacy and confidentiality were safeguarded through coded data, secure storage, and restricted access, with no identifying information included in published results.

3. Results

3.1. Effectiveness of Computer-Aided Assessment (CAA) in Enhancing Students' Critical Thinking Skills through Non-routine Problem-Solving

This study examined the effectiveness of Computer-Aided Assessment (CAA) in enhancing students' critical thinking skills through non-routine problem solving. CAA provided interactive non-routine problem-solving activities and instant verification feedback, which aimed to help students enhance their critical thinking skills. To measure the effectiveness of the CAA, a comparison of the pretest and post-test for the experimental group was conducted. A comparison was also made between the post-test scores of the experimental and control groups.

The following tables show the results of the statistical analysis. Table 2a presents the results of the paired t-test conducted to compare the pretest and post-test mean scores of critical thinking skills in the experimental group, aiming to determine whether significant differences exist between them.

Table 2a. Paired t-test Comparing Pretest and Post-test Mean Scores of Critical Thinking Skills of the Experimental Group

Learning Delivery Modalities	Assessment Measures	Descriptive Statistics		t-value	p-value
Balanced blended	Pretest	Mean	5.24	-9.457	p = 0.000 < 0.05
		S.D.	3.08		
		Verbal Interpretation	Less Critical		
	Post-test	Mean	10.00		
		S.D.	3.17		
		Verbal Interpretation	Critical		
Predominantly F2F blended	Pretest	Mean	6.74	-8.302	p = 0.000 < 0.05
		S.D.	3.14		
		Verbal Interpretation	Sufficiently Critical		
	Post-test	Mean	11.76		
		S.D.	2.91		
		Verbal Interpretation	Critical		
Fully online	Pretest	Mean	4.20	-4.987	p = 0.000 < 0.05
		S.D.	2.83		
		Verbal Interpretation	Less Critical		
	Post-test	Mean	6.92		
		S.D.	2.36		
		Verbal Interpretation	Sufficiently Critical		

As shown in Table 2a, significant differences were found between the pretest and post-test mean scores of critical thinking skills in the experimental group (EG) across the three learning modalities (p = 0.000), indicating a significant improvement from pretest to post-test. In the balanced blended modality, the EG improved from 'Less Critical' (Mean = 5.24, SD = 3.08) to 'Critical' (Mean = 10.00, SD = 3.17). In the pre-dominantly blended modality, the EG improved from 'Sufficiently Critical' (Mean = 6.74, SD = 3.14) to 'Critical' (Mean = 11.76, SD = 2.91). In the fully online modality, the EG improved from 'Sufficiently Critical' (Mean = 4.20, SD = 2.83) to 'Critical' (Mean = 6.92, SD = 2.36).

Table 2b presents the results of the independent t-test comparing the post-test mean scores of critical thinking skills between the experimental and control groups to determine whether differences exist and whether they are statistically significant.

Table 2b. Independent t-test Comparing Post-test Mean Scores of Critical Thinking Skills Between Experimental and Control Groups

Learning Delivery Modalities	Groups	Descriptive Statistics		t-value	p-value
Balanced blended	Experimental	Mean	10.00	1.228	p = 0.224 > 0.05
		S.D.	3.17		
		Verbal Interpretation	Critical		
	Control	Mean	8.94		
		S.D.	4.16		

		Verbal Interpretation	Sufficiently Critical		
Pre-dominantly F2F blended	Experimental	Mean	11.76	5.413	p = 0.000 < 0.05
		S.D.	2.91		
		Verbal Interpretation	Critical		
	Control	Mean	6.66		
		S.D.	3.44		
		Verbal Interpretation	Sufficiently Critical		
Fully online	Experimental	Mean	6.92	-2.025	p = 0.048 < 0.05
		S.D.	2.36		
		Verbal Interpretation	Sufficiently Critical		
	Control	Mean	8.47		
		S.D.	3.38		
		Verbal Interpretation	Sufficiently Critical		

As shown in Table 2b, no significant difference was found between the experimental group (EG) and the control group (CG) in the balanced blended modality (t = 1.228, p = 0.224). However, significant differences were observed between the EG and CG in both the predominantly face-to-face (F2F) blended modality (t = 5.413, p = 0.000) and the fully online modality (t = -2.025, p = 0.048). In the predominantly F2F blended modality, the EG (Mean = 11.76, SD = 2.91), categorized as 'Critical,' outperformed the CG (Mean = 6.66, SD = 3.44), categorized as 'Sufficiently Critical.' Interestingly, in the fully online modality, the CG (Mean = 8.47, SD = 3.38) outperformed the EG (Mean = 6.92, SD = 2.36), though both groups still fell within the 'Sufficiently Critical' level.

3.2. Differences in the Impact of Computer-Aided Assessment (CAA) on Critical Thinking Skills Across Different Learning Delivery Modalities

This study also examined the effect of Computer-Aided Assessment (CAA) on students' critical thinking skills across different learning delivery modalities. Balanced blended, predominantly F2F blended, and fully online modalities may have influenced students' ability to engage in critical thinking and non-routine problem-solving differently when using CAA. To investigate these differences, a two-way ANOVA was conducted to analyze the impact of CAA on post-test critical thinking skills mean scores across various learning delivery modalities.

Table 3 presents the results of the two-way ANOVA, which analyzed the interaction between CAA implementation and the learning delivery modality, examining critical thinking skills across three modalities—balanced blended, predominantly blended, and fully online—in both the experimental and control groups based on their post-test scores.

Table 3. Two-Way ANOVA for Impact of CAA on Post-test Critical Thinking Skills Mean Scores Across Different Learning Delivery Modalities

Source of Variation	F-value	p-value	Significance	Remarks (Post hoc)
Group (Experimental/Control)	8.870	p = 0.003 < 0.05	Significant	Balanced blended VS Fully online
Learning Delivery Modality	4.295	p = 0.015 < 0.05	Significant	
Interaction (Group*Learning Delivery Modality)	12.775	p = 0.000 < 0.05	Significant	

As shown in Table 3, significant differences were observed between groups (F = 8.870, p = 0.003) and among the learning delivery modalities (F = 4.295, p = 0.015). Additionally, significant interaction effects were observed between groups and learning delivery modalities (F = 12.775, p = 0.000).

The significant differences between groups were found in both the predominantly blended and fully online modalities, as shown in Table 2b. Significant differences among learning delivery modalities were further identified through post hoc tests, which revealed differences between the balanced blended and fully online modalities. The experimental group (EG) in the balanced blended modality, considered at the 'Critical' level, outperformed those in the fully online modality, who were considered at the 'Sufficiently Critical' level, as shown in Table 2b (Mean = 10.00, SD = 3.17; Mean = 6.92, SD = 2.36, respectively).

4. Discussion

The results showed that participants in all modalities showed progress, with both blended modalities reaching the 'Critical' level and even the lowest-performing modality maintaining a 'Sufficiently Critical' level. Standard deviations indicated varying consistency in performance, with lower values reflecting more stable improvements. These findings highlight the need for tailored instructional approaches to enhance critical thinking skills across different learning delivery modalities, ensuring that all students achieve higher levels of critical thinking proficiency. These results align with studies that reported students' "above average" critical thinking skills (Alcantara & Bacsa, 2017; Kuşcu & Erdoğan, 2024). Conversely, some studies found that students' critical thinking skills ranged from "fairly low" to "moderate" levels (Fadhlullah & Ahmad, 2017; Hasanah et al., 2019).

Building on these findings, this study further demonstrated the effectiveness of computer-aided assessment (CAA) in enhancing students' critical thinking skills through non-routine problem-solving. The paired t-test results demonstrated significant improvements in critical thinking skills from pretest to post-test across all three learning delivery modalities in the experimental group (EG). These results showed that technology-enhanced or digital tools, such as CAA, enhance students' critical thinking skills, aligning with the findings of Pramasdyahsari et al. (2023), who found that the implementation of the digital book STEM-PjBL (STEM-Project-Based Learning) led to a significant gain in fostering students' critical thinking skills. Similarly, Santos and Bastos (2021) found that digital tools enhanced students' critical thinking skills by increasing engagement with the material and providing immediate feedback.

The independent t-test results showed differences in CAA's impact across learning delivery modalities. In the balanced blended modality, no significant difference was found between the experimental and control groups. However, in the predominantly face-to-face (F2F) blended modality, the EG showed significantly greater development compared to the control group (CG). Interestingly, in the fully online modality, the CG outperformed the EG, though both remained within the 'Sufficiently Critical' level. These findings suggested that CAA's effectiveness varied by learning delivery modality.

Specifically, the balanced blended modality supported critical thinking equally in both groups, while the predominantly F2F blended modality benefited more from CAA integration. In contrast, the fully online control group outperformed the experimental group. This unexpected result, in which the fully online CG outperformed the EG using CAA, may have been influenced by several factors. The EG may have struggled to adapt to CAA, particularly if they lacked prior experience with such assessment methods or required additional guidance. Additionally, they may have faced technical challenges or experienced an increased cognitive load. External factors such as home distractions or limited internet access, may have further affected engagement and performance. Furthermore, CAA may have altered instructor and peer interactions, reducing the support available to the EG. In contrast, the CG, having relied on familiar assessment methods, may have exhibited more consistent performance. This unexpected result raises questions about students' readiness for self-directed learning, the quality of CAA implementation, and the need for enhanced support mechanisms.

These findings highlight the need to carefully integrate CAA into fully online learning environments. The unexpected results raise important questions about students' readiness for self-directed learning, the quality of CAA implementation, and the necessity of enhanced support mechanisms. Ultimately, these results underscore the importance of tailoring instructional strategies to each modality for optimal learning outcomes. This findings aligned with Mahanal et al. (2019), whose study on the RICOSRE problem-based learning model similarly found improved critical thinking skills in students within the blended modality.

The results of the two-way ANOVA further emphasized CAA's varying impact across different modalities, showing significant main effects and a notable interaction effect. Post hoc tests revealed that the balanced blended modality consistently outperformed the fully online modality in enhancing critical thinking skills, with lower standard deviations indicating more uniform improvement. Among the modalities, the balanced blended modality emerged as the most effective learning environment for CAA intervention. These findings aligned with the work of Suana et al. (2020), who advocated for mixed approaches combining face-to-face instruction with online learning to improve critical thinking skills. Similarly, Simonovic et al. (2022) found that interactive and immediate feedback mechanisms in brief online workshops enhanced students' critical thinking, suggesting that incorporating workshop-style activities into CAA could strengthen its effectiveness. Furthermore, Sako (2024) demonstrated that AI-assisted task-based collaborative learning improved critical thinking, digital literacy, and creativity, although its success depended on proper implementation and instructor support.

5. Conclusion

The study revealed that computer-aided assessment (CAA) significantly enhanced students' critical thinking skills through non-routine problem solving, with significant improvements observed across all three learning delivery modalities. However, the impact of CAA varied across these learning delivery modalities: the predominantly face-to-face (F2F) blended modality showed the greatest improvement in the experimental group, while the fully online modality experienced better performance in the control group. Among the three, the balanced blended modality benefited the most from the CAA intervention and emerged as the most effective overall, exhibiting the highest and most consistent gains in critical thinking skills. These findings highlight the importance of understanding how different modalities shape the impact of CAA on critical thinking development and underscore the need to tailor instructional strategies to specific learning environments to maximize its effectiveness.

References

Abulhul, Z. (2021). Teaching Strategies for Enhancing Student's Learning. *Journal of Practical Studies in Education, 2(3), 1-4.* https://doi.org/10.46809/jpse.v2i3.22

Aini, N.R., Syafril, S., Netriwati, N,, Rahayu, T., & Puspasari. (2019). *IOP Conference Series Journal of Physics: Conf. Series 1155(2019) 0120*26. https://doi.org/10.1088/1742-6596/1155/1/012026

Ajogbeje, O.J. (2023). Enhancing Classroom Learning Outcomes: The Power Of Immediate Feedback Strategy. *International Journal of Disabilities Sports and Health Sciences, 6(3), 453-465*. https://doi.org/10.33438/ijdshs.1323080

Alcantara, E. C., & Bacsa, J. M. P. (2017). Critical Thinking and Problem Solving Skills in Mathematics of Grade-7 Public Secondary Students. *Asia Pacific Journal of Multidisciplinary Research, 5(4), 21-27*. https://www.apjmr.com/wp-content/uploads/2017/12/APJMR-2017.5.4.2.03.pdf

Andrade, R.R., Fortes, E.C., & Mabilangan, R.A. (2020). Problem Solving Heuristics and Mathematical Abilities of Heterogeneous Learners. *Universal Journal of Educational Research 8(11), 5114-5126*. https://doi.org/10.13189/ujer.2020.081111

Anstey, L., & Watson, G. (2018). Rubric for eLearning Tool Evaluation. Centre for Teaching and Learning Western University. https://teaching.uwo.ca/pdf/elearning/Rubric-for-eLearning-Tool-Evaluation.pdf

Barana, A., Conte, A., Fioravera, M., Marchisio, M., & Rabellino, S. (2018). A model of formative automatic assessment and interactive feedback for STEM. 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC). https://doi.org/10.1109/COMPSAC.2018.00178

Barana, A., Marchisio, M., & Sacchet, M. (2021). Interactive feedback for learning mathematics in a digital learning environment. Education Sciences, 11(6), 279. https://doi.org/10.3390/educsci110 60279

Berisha, V. (2015). The General Characteristics Of Mathematics Textbooks For Lower Secondary School In Kosovo. *International Journal of Novel Research in Education and Learning, 2(2), 19-23*. https://www.noveltyjournals.com

Berestova, A., Kolosov, S., Tsvetkova, M., & Grib, E. (2022). Academic motivation as a predictor of the development of critical thinking in students. *Journal of Applied Research in Higher Education, 14(3), 1041-1054*. https://doi.org/10.1108/JARHE-02-2021-0081

Demirel, M., Derman, I., & Aran, O. C. (2017). Examining Graduate Dissertations in the Field of Critical Thinking: A Case from Turkey. *Eurasian Journal of Educational Research, 67(55), 141–159*. http://dx.doi.org/10.14689/ejer.2017.67.9

Dwi Susandi, A., Sa'Dijah, C., Rahman As'Ari, A., & Susiswo. (2019). Students' critical ability in mathematics is based on cognitive styles. *Journal of Physics: Conference Series, 1315(1)*. https://doi.org/10.1088/1742 6596/1315/1/012018.

Fadhlullah, A. & Ahmad, N. (2017). Thinking Outside of the Box: Determining Students' Level of Critical Thinking Skills in Teaching and Learning. *Asian Journal of University Education, 51-70*. https://files.eric.ed.gov/fulltext/EJ1207765.pdf

Fan, L. & Zhu, Y. (2000). Problem Solving in Singaporean Secondary Mathematics Textbooks. *The Mathematics Educator, 5(1), 117-141*. https://repository.nie.edu.sg/server/api/core/bitstreams/978a61aa-a672-4f1a-ba95-605fdb26d399/content

Ghazali, M.A.I.M, Ali, Z., & Rahim, A.A.A. (2022). Adopting Design and Development Research (DDR) Approach in Framing the ESL Preservice Teachers' Roles and Responsibilities to Teach Writing Skills. *Journal of Pharmaceutical Negative Results, 13(6), 695-704*. https://doi.org/10.47750/pnr.2022.13.S06.098

Gavaz, H. O., Yazgan, Y., & Arslan, C. (2021). Non-routine problem solving and strategy flexibility: A quasi-experimental study. *Journal of Pedagogical Research, 5(3), 40-54*. https://doi.org/10.33902/JPR.2021370581

Hafni, R. N., Herman, T., Nurlaelah, E., & Mustikasari, L. (2020). The importance of science, technology, engineering, and mathematics (STEM) education to enhance students' critical thinking skills in facing Industry 4.0. *Journal of Physics: Conference Series, 1521(4), 0–7*. https://doi.org/10.1088/1742-6596/1521/4/042040.

Hasanah, S. N., Sunarno, W,. & Prayitno, B. A. (2019). Profile of Students' Critical Thinking Skills in Junior High Schools in Surakarta. 3rd International Conference on Learning Innovation and Quality Education (ICLIQE 2019), Advances in Social Science, Education and Humanities Research, 397, 570-575. https://doi.org/10.2991/assehr.k.200129.070

Hujjatusnaini, N., Corebima, A., Prawiro, S., & Gofur, A. (2022). The effect of blended project-based learning integrated with 21st-century skills on pre-service biology teachers' higher-order thinking skills. *Jurnal Pendidikan IPA Indonesia, 11(1), 104-118*. https://doi.org/10.15294/jpii.v1.1i1.27148.

Jaya, S. et al. (2021) Employing Design and Development Research (DDR) Approach in Designing Next Generation Learning Spaces (NGLS) In Teachers' Pedagogy and Technology Tools. *Review of International Geographical Education (RIGEO), 11(7), 1237-1246*. https://doi.org/10.48047/rigeo.11.07.116

Jhangiani, R.S., Chiang, I. A., Cuttler, C., & Leighton, D.C. (2019). Research methods in psychology (4th ed.). Kwantlen Polytechnic University. https://kpu.pressbooks.pub/psychmethods4e/

Julie, H., Sanjaya, F., & Anggoro, A.Y. (2017). The students' ability in the mathematical literacy for uncertainty problems on the PISA adaptation test. *AIP Conference Proceedings*. https://doi.org/10.1063/1.4995153

Kablan, Z., & Uğur, S. S. (2019). The relationship between routine and non-routine problem solving and learning styles. *Educational Studies, 1-16*. https://doi.org/110.1080/03055698.2019.1701993

Kolovou, A., van den Heuvel-Panhuizen, M., & Bakker, A. (2009). Non-Routine Problem Solving Tasks in Primary School Mathematics Textbooks – A Needle in a Haystack. *Mediterranean Journal for Research in Mathematics Education, 8 (2), 31-68*. https://www.researchgate.net/publication/233864424

Kundu, A., & Bej, T. (2020). Experiencing e-assessment during COVID-19: An analysis of Indian students' perception. *Higher Education Evaluation and Development, 15(2), 114-134*. https://doi.org/10.1108/heed 03-2021-0032

Kuşcu, H., & Erdoğan, F. (2024). Examination of Creative Thinking Skills and Critical Thinking Dispositions of Pre Service Mathematics Teachers. *International Journal of Progressive Education, 20(5), 38-55*. https://doi.org/10.29329/ijpe.2024.1063.4

Mahanal, S., Zubaidah, S., Sumiati, I. D., Sari, T. M., & Ismirawati, N. (2019). RICOSRE: A learning model to develop critical thinking skills for students with different academic abilities. *International Journal of Instruction, 12(2), 417-434*. https://doi.org/10.29333/iji.2019.12227a.

Manopo & Lisarani, V. (2021). A Comparative Analysis of Non-Routine Problems in Mathematics Textbooks of Indonesia and Singapore. *EDUMATIKA: Jurnal Riset Pendidikan Matematika, 4(1)*. https://doi.org/10.32939/ejrpm.v4i1.750

Natarajan, D., Al Sulaimi, B.H., Lacap. E.M. & kumar, K. (2022). A Comparative Study of Student Mathematical Performance in three modes of Teaching and Learning Education during the Covid-19 Pandemic. *International Journal of Current Education Studies (IJCES), 1(1), 36-50*. https://doi.org/10.5281/ zenodo.6762580

Noh, S.C., & Karim, A.M.A. (2021). Design thinking mindset to enhance education 4.0 competitiveness in Malaysia. *International journal of evaluation and research in education 10 (2),494-501*. https://doi.org/10.11591/ ijere.v10i2.20988

Padzil, M.R., Karim, A.A., & Husnin, H. (2021). Employing DDR to Design and Develop a Flipped Classroom and Project based Learning Module to Applying Design Thinking in Design and Technology. *(IJACSA) International Journal of Advanced Computer Science and Applications, 12(9), 791-798*. https://doi.org/10.14569/IJACSA.2021.0120988

Pramasdyahasari, A.S. et al. (2023). Fostering students' mathematical critical thinking skills on number patterns through digital book STEM PjBL. *EURASIA Journal of Mathematics, Science and Technology Education, 2023, 19(7), em2297*. https://doi.org/10.29333/ejmste/13342

Rahmasari, S.M. Prabowo, A., Zaenuri, Z., & Waluya, B. (2023). Improving students' mathematical critical thinking abilities: A systematic literature review. *Contemporary Educational Researches Journal. 13(4), 288-305.* https://doi.org/10.18844/cerj.v14i4.9257

Rural, J.D., Amora, E.E., De Boraj, R.J.I., De La Torre, J.A., Duenas, P.T., & Jardin Jr., E.S. (2022). Mathematics Academic Performance: It's Relationship to Attitudes and Engagement in Online Learning. European *Online Journal of Natural and Social Sciences 2022; Special Issue on Education Research in COVID Era: The Future of the Next Generation, 11(1), 42-53*. https://european-science.com/eojnss_proc/article/viewFile/6411/2902

Sako T. (2024). Enhancing Critical Thinking through AI-Assisted Collaborative Task-Based Learning. *Journal of English Language Teaching and Linguistics, 9(2), August 2024, 157-170*. http://dx.doi.org/10.21462/jeltl.v9i2.1319

Santos, V., & Bastos, N. R. O. (2021). Critical thinking on mathematics in higher education: Two experiences [Conference session]. *Conference on Communications in Computer and Information Science*. https://doi.org/10.1007/978-3-030-73988-11

Scriven, M.., & Paul, R. (2005). The critical thinking community. https://www.criticalthinking.org/pages/defining-critical-thinking/766

Simonovic, B., Vione, K.C., Fido, D., Stupple, E. J. N., Martin, J., & Clarke, R. (2022). The impact of attitudes, beliefs, and cognitive reflection on the development of critical thinking skills in online students. *Online Learning, 26(2), 254-274*. https://doi.org/10.24059/olj.v26i2.2725

Suana, W., Istiana, P., &; Maharta, N. (2019). The effect of the application of blended learning on static electricity material on students' critical thinking skills. *Journal of Science Education (Jps), 7(2), 129*. https://doi.org/10.26714/jps.7.2.2019.129-136

Tanu Wijaya, T. (2020). How chinese students learn mathematics during the coronavirus pandemic. IJERI: International Journal of Educational Research and Innovation, 15, 1 16. https://doi.org/10.46661/ijeri.4950

Tong, D.H., Uyen, B.P., & Ngan, L.K. (2022). The effectiveness of blended learning on students' academic achievement, self-study skills and learning attitudes: A quasi-experiment study in teaching the conventions for coordinates in the plane. Heliyon 8 (2022) e12657, 1-14. https://doi.org/10.1016/j.heliyon.2022.e12657

van Zanten, M. & van den Heuvel-Panhuizen, M. (2018). Opportunity to learn problem solving in Dutch primary school mathematics textbooks. *ZDM*. https://doi.org/10.1007/s11858-018-0973-x

Wang, Q., & Abdullah, A.H. (2024). Enhancing Students' Critical Thinking Through Mathematics in Higher Education: A Systemic Review. *SAGE Open, July-September 2024: 1-15*. https://doi.org/10.1177/21582440241275651

Yumiati, Y., & Kusumah, Y. S. (2019). Interaction between Students' Learning and Early Mathematical Skills to Increase Mathematical Critical Thinking Skills. Al-Jabar*: Jurnal Pendidikan Matematika, 10(1), 125-134*. https://doi.org/10.24042/ajpm.v10i1.3712.

Zukhairina, Wujiati, Sari, S. R., Yennizar, Zulqarnain, &, & Pujiarto, P. (2020). Parenting practices of early childhood education teachers in developing early childhood communication and critical thinking skills: Case studies. *Journal of Critical Reviews, 7(7), 125–132*. https://doi.org/10.31838/jcr.07.07.20