

Digital Self-Assessment as a Tool for Academic Improvement among Female Students: Evidence from an Intermediate College in Karachi

Muhammad Zeeshan Rub*

Abstract

The integration of information technology (IT) into assessment has become increasingly important in contemporary pedagogy; however, its effectiveness within South Asian intermediate institutions particularly girls-only colleges remains insufficiently explored. This study examined whether digital self-assessment quizzes enhance academic performance among female intermediate students compared to traditional paper-based tests. Using a quasi-experimental repeated-measures design, 80 students were randomly assigned to either a control group (traditional assessments) or an experimental group (digital quizzes). Both assessments targeted identical learning outcomes, and a pilot test established acceptable reliability (Cronbach's $\alpha = .78$). A mixed ANOVA revealed a significant main effect of Time, $F(1, 78) = 52.34, p < .001$, partial $\eta^2 = .402$, indicating improvement from Test 1 to Test 2 across groups. A significant main effect of Group, $F(1, 78) = 13.26, p < .001$, partial $\eta^2 = .145$, showed that students using digital quizzes performed better overall. However, the nonsignificant Time $\times$ Group interaction proposes that digital quizzes did not result comparatively greater short-term learning improvements than traditional assessments. These findings indicate that digital self-assessments support improved performance, and their impact may be moderated by contextual barriers like digital literacy gaps, unequal access to devices, and infrastructural limitations which are more common in South Asian educational settings. By strengthening the institutional capacity, enhancing digital literacy, and adopting long-term implementation strategies, the pedagogical value of digital assessments may optimize.

Keywords: Digital Self-Assessment, Student Learning Outcomes, Mixed ANOVA, South Asian Education

* Independent Researcher from Karachi, Pakistan. ORCID iD: <https://orcid.org/0009-0002-1673-8205>, Email: zeeshan4153@qureshi@gmail.com

Introduction

The integration of information technology (IT) in educational assessment has become a defining feature of modern teaching and learning in the fast-paced world. Digital self-assessment tools, such as online quizzes, interactive formative tests, and automated feedback systems not only offer opportunities to enhance learner engagement, provide immediate feedback, but also promote self-regulated learning. The need to evaluate the effectiveness of digital assessments across diverse contexts has become increasingly important as global education systems continue to transition toward technology-enabled instruction.

Many intermediate and secondary institutions in Pakistan continue to rely predominantly on traditional, paper-based assessments despite the worldwide shift toward digital learning environments. Such practices often limit timely feedback, constrain student engagement, and restrict opportunities for personalized learning pathways. Although international studies highlight the potential advantages of digital self-assessments, including increased motivation and improved learning outcomes, their applicability within South Asian educational settings remains under-examined. This gap is especially evident in girls-only institutions, where cultural norms, unequal access to technology, and varied levels of digital literacy may influence the integration and effectiveness of digital assessment tools.

Given these contextual challenges, there is a persistent need to generate localized empirical evidence on how digital self-assessment practices influence learning outcomes among intermediate-level female students. Understanding whether digital assessment tools meaningfully enhance performance beyond the benefits already attributed to structure testing can guide educators, school leaders, and policymakers in designing technology-supported assessment systems that are both practical and equitable.

Research Question

Does the implementation of digital self-assessment quizzes improve student performance compared to traditional assessments?

Objective

The objective of this study is to evaluate the effectiveness of digital self-assessment quizzes in improving academic performance among female intermediate students and to compare learning outcomes between digital and traditional assessment modalities.

LITERATURE REVIEW

Enhancing Engagement and Interactivity

Research consistently emphasizes that digital assessments can increase student engagement and interactivity. Digital quizzes, multimedia elements, and instant response mechanisms transform passive learners into active participants, supporting cognitive, emotional, and behavioral engagement (Godsk & Møller, 2024). Recent studies also indicate that gamified elements further boost motivation and participation (Hamari et al., 2016). However, poorly designed digital assessments may lead to superficial learning or distraction.

Contradictory Evidence: Tech Fatigue and Cognitive Overload

While many studies emphasize increased engagement with digital assessments, contradictory findings also exist. Excessive screen time and poorly structured quizzes can cause tech fatigue, reducing motivation and concentration (Saleem, Chikhaoui, & Malik, 2024). Additionally, multimedia-rich quizzes may lead to cognitive overload, hindering deep learning and causing students to rely on guessing rather than actual understanding (Sweller, 2020). These contradictions highlight the importance of balancing interactivity with cognitive simplicity.

Immediate Feedback and Learning Outcomes

Delivery of immediate feedback is a major strength of digital assessments. Students benefit from timely feedback, they identify misconceptions, adjust strategies, and engage in self-regulated learning (Black & Wiliam, 1998; Shute, 2008). Recent research demonstrates that prompt feedback enhances retention and test performance (Huang, Stephens, & Brown, 2025). Comparatively, generic automated feedback may be less effective for deeper conceptual learning.

Adaptive Learning and Personalization

Adaptive assessments adapt difficulty based on student performance, supporting differentiated instruction (Zhang et al., 2004). According to Tomlinson (2014), personalized pathways enhance motivation and self-efficacy. Nevertheless, adaptive systems require large item banks and sufficient digital literacy, which are always challenging for under-resourced schools. Recent studies highlight that limited technological infrastructure and gaps in digital skills continue to hinder effective implementation of technology-enabled assessment tools in such contexts (Nikou & Aavakare, 2021; Simon & Zeng, 2024).

Efficiency, Accessibility, and Digital Readiness

According to García-Peñalvo et al. (2020), digital assessments streamline grading, support remote learning, and enhance digital literacy. However, inequalities in access to devices and internet connectivity pose significant challenges.

South Asian Contextual Barriers

In the South Asian context, challenges including limited device availability, unstable internet connectivity, and varying digital literacy levels hinder effective implementation of digital assessments (Waqar et al., 2024). Cultural and gender-related factors also influence access to technology. In girls-only colleges, socio-cultural restraints may restrict access to digital tools at home, reducing consistency in digital learning experiences. These barriers highlight the importance of context-sensitive research and localized implementation strategies.

Synthesis and Gaps

The literature indicates that digital self-assessments improve engagement, feedback, and personalization. However, significant gaps remain, particularly in intermediate-level and South Asian contexts. Few studies examine digital assessment challenges such as cognitive overload or tech fatigue, and even fewer evaluate traditional vs. digital assessments within the same cohort.

These gaps demonstrate the need to examine digital self-assessment within an intermediate girls' college in Pakistan.

Methodology

This study employed a quasi-experimental mixed factorial design (2 x 2 Mixed ANOVA), where Group (Control vs. Experimental) was the between-subjects factor, and Time (Time 1 vs. Time 2) was the within-subjects factor, to investigate the impact of digital self-assessment quizzes on student performance in an intermediate college section. The study was conducted over a period of five months in a girls-only college. A total of 80 intermediate-level female students were randomly selected from the eligible population and subsequently randomly assigned to one of the two groups ($N=40$ each), the control group or the experimental group. Each group was assessed twice at two time points: an initial test (Time 1) and a follow-up test (Time 2) aligned with the same learning outcomes. This design thus allowed for the evaluation of within-subject improvements over time (Time 1 $\rightarrow$ Time 2) and between-group differences (traditional vs. digital assessment).

The control group was assessed using traditional paper-based methods, i.e., hard-copy assessment items administered and graded manually. The experimental group, however, utilized digital quizzes hosted on the Microsoft Forms platform. These quizzes were intentionally designed to be interactive and engaging, incorporating multiple-choice questions. The key intervention feature was immediate automated feedback, delivered as automated scoring and corrective explanations immediately following quiz submission. The teacher, co-teacher and ICT-assistant of the college were responsible for planning, conducting, and supervising all assessments. Essential resources for the digital intervention included devices with stable internet connectivity, secure browsers, and technical support to ensure smooth administration of the quizzes.

Both groups were assessed using two versions of the assessment items, corresponding to the two time points. The tests (Time 1 and Time 2) were developed to be aligned with the same learning objectives. Test 1 was set at medium difficulty, while Test 2 was set at slightly higher difficulty to mitigate practice effects. The strategic difference was the mode of administration: the control group took both Time 1 and Time 2 assessments using paper-based tests, while the experimental group took both Time 1 and Time 2 assessments using digital quizzes.

Difficulty levels were reviewed by senior faculty to ensure equivalence in curricular coverage and cognitive demand.

Reliability, Validity, and Ethical Considerations

The assessment items were reviewed for content validity by the college faculty to ensure alignment with the curriculum. The study received approval from the college management, and written informed consent was obtained from all participants. No external funding was involved, and the author declares no conflicts of interest. To ensure reliability, a pilot quiz was administered to a separate group of students not included in the final study. Cronbach's alpha was calculated and yielded a reliability coefficient of 0.78, indicating acceptable internal consistency for the quiz items.

Results

Table 1.

Means and Standard Deviations for Test Scores by Group (N = 40 per group)

	Group	Mean	Std. Deviation	N
Test 1	Control	11.10	1.44	40
	Experimental	12.20	1.41	40
Test 2	Control	12.10	1.50	40
	Experimental	13.10	1.29	40

Descriptive statistics in Table 1 showed that both groups improved from Test 1 to Test 2. In the control group, mean scores increased from 11.10 (SD = 1.44) to 12.10 (SD = 1.50); in the experimental group, mean scores increased from 12.20 (SD = 1.41) to 13.10 (SD = 1.29).

Table 2.

Estimated Marginal Means (With 95% Confidence Intervals)

Factor	Level	Mean	SE	CI
Time	Test 1	11.65	.165	11.32- 11.98
Time	Test 2	12.60	.151	12.30- 12.90
Group	Control	11.60	.204	11.19- 12.00
Group	Experimental	12.65	.204	12.24- 13.06

In Table 2 the estimated marginal means show that overall performance increased from Test 1 ($M = 11.65$, 95% CI [11.32, 11.98]) to Test 2 ($M = 12.60$, 95% CI [12.30, 12.90]), indicating a significant improvement across time. The confidence intervals for the two time points do not overlap, further supporting this improvement. For the between-subjects factor, the experimental group showed a higher overall mean score ($M = 12.65$, 95% CI [12.24, 13.06]), compared to the control group ($M = 11.60$, 95% CI [11.19, 12.00]), indicating that students who took digital quizzes performed better overall. However, these differences do not speak to improvement over time; they simply show group-level differences across both assessments.

Table 3.

Mixed ANOVA: Tests of Within-Subjects Effects (Time and Time $\times$ Group)

Effect	SS	Df	MS	F	P	Partial η^2
Time	36.10	1	36.10	52.34	<.001	.402
Time $\times$ Group	0.10	1	0.10	0.15	.704	.002
Error (Time)	53.80	78	0.69	---	---	---

Table 3 presents result of the within-subjects ANOVA. There is a statistically significant effect of Time, $F(1, 78) = 52.34$, $p < .001$, partial $\eta^2 = .402$, indicating a strong improvement from Test 1 to Test 2. However, the Time $\times$ Group interaction is not significant, $F(1, 78) = 0.15$, $p = .704$, partial $\eta^2 = .002$, suggesting that the rate of improvement does not differ between the control and experimental groups. Both groups improved, but the digital assessment did not produce a significantly larger gain than the traditional method.

Table 4.
Between-Subjects Effects (Group Differences)

Effect	SS	Df	MS	F	P	Partial η^2
Group	44.10	1	44.10	13.26	<.001	.145
Error	259.40	78	3.33	---	---	---

The between-subjects analysis shows a significant main effect of Group, $F(1, 78) = 13.26$, $p < .001$, partial $\eta^2 = .145$, indicating that the experimental group scored higher overall than the control group across both time points. This reflects a moderate effect size, suggesting that students exposed to digital self-assessment methods generally performed better. However, this effect does not indicate that they improved more over time only that their overall scores were higher.

Table 5.
Pairwise Comparisons for Time (Bonferroni Corrected)

Contrast	Mean Diff	SE	P	95% CI
Test 1 $\rightarrow$ Test 2	-0.95*	.131	<.001	-1.21—-0.69

Note. The mean difference is significant at $p < .05$

Pairwise comparisons reveal a significant difference between Test 1 and Test 2 scores, with students scoring on average 0.95 points higher on Test 2 than on Test 1 ($p < .001$). The 95% confidence interval for this difference (-1.21 to -0.69) confirms the reliability of this improvement. This finding demonstrates that performance increased significantly over time, consistent with the main effect of Time shown in the mixed ANOVA.

Table 6.
Means for Each Group across Both Time Points

Group	Test 1 Mean	Test 2 Mean	Improvement
Control	11.10	12.10	+1.00
Experimental	12.20	13.10	+0.90

Table 6 shows that the control group improved by +1.00 points, whereas the experimental group improved by +0.90 points. The improvement trends are nearly identical, which explains the nonsignificant Time $\times$ Group interaction in Table 3. Both groups benefited from repeated assessment, but the digital quizzes did not produce a disproportionately higher improvement compared to traditional paper-based tests. This supports the conclusion that digital assessments lead to higher overall performance levels, they do not necessarily produce greater short-term learning gains.

Discussion

The purpose of this study was to evaluate the effectiveness of digital assessment quizzes in improving academic performance among intermediate-level female students. The findings indicate that while both the control (traditional) and experimental (digital) groups demonstrated significant improvements from Time 1 to Time 2, the rate of improvement did not differ between groups.

Interpretation of Core Findings

The significant main effect of Time ($F(1, 78) = 52.34, p < .001$, partial $\eta^2 = .402$) indicates that repeated testing led to measurable gains in performance, consistent with literature on formative assessment and feedback (Shute, 2008; Black & Wiliam, 1998). This improvement occurred irrespective of the mode of formative assessment, highlighting the value of structured testing in supporting student learning.

The significant main effect of Group ($F(1, 78) = 13.26, p < .001$, partial $\eta^2 = .145$) reflects that the experimental group had higher overall scores across both time points. However, the non-significant Time $\times$ Group interaction ($F(1, 78) = 0.15, p = .704$, partial $\eta^2 = .002$) indicates that the digital quizzes did not accelerate learning more than traditional assessments. Thus, while the experimental group performed slightly better on average, this difference should not be interpreted as evidence of longer learning gains attributable to digital self-assessment.

Contextual Considerations

Both assessment modes were aligned with the same learning objectives, and the experimental group's digital quizzes included interactive features and instant feedback. The absence of a differential improvement effect may be influenced by factors such as the limited duration of exposure, relatively simple quiz design to avoid cognitive overload, and controlled classroom environments that minimized differences in engagement opportunities. Additionally, regional constraints like varying digital literacy and limited prior exposure to digital platforms may have restricted potential gains from the digital intervention (Jamil & Muschert, 2023).

Comparison with Previous Research

The results partially align with research demonstrating motivational gains and improved accuracy via digital quizzes (Huang, Stephens, & Brown, 2025). However, the absence of a Time $\times$ Group interaction

contrasts with studies indicating that digital assessments can produce stronger or equivalent learning gains compared to traditional tests (Zheng, Bender, & Lyon, 2021). This discrepancy may be attributed to infrastructural limitations, novelty effects, and variations in instructional design common in developing educational contexts.

Limitations

Several limitations must be acknowledged:

1. Sample size and scope only 80 students from a single girls-only college limits generalizability.
2. Limited item bank: restricted the adaptive or diagnostic power of digital quizzes.
3. Practice effects: repeated-measures design may have inflated performance improvements despite slight difficulty adjustments.

Pedagogical Implications

Despite the lack of a differential effect on learning improvement, digital quizzes provide pedagogical benefits beyond immediate score gains. Features like instant feedback, automated scoring, and interactive engagement support self-regulated learning, enhance motivation, and build students' familiarity with technology (Hamari et al., 2016). For South Asian intermediate colleges, this study demonstrates that integrating digital assessments is feasible and can enhance overall learning experiences, even if short-term performance gains are comparable to traditional assessments.

Conclusion

This study shows that digital self-assessment quizzes are associated with higher overall scores but do not produce significantly greater short-term learning gains compared to traditional paper-based assessments. The critical finding is that both groups improved similarly over time, indicating that repeated assessment itself, rather than assessment mode, was the primary driver of improvement.

Nevertheless, digital quizzes offer important pedagogical advantages, including timely feedback, learner autonomy, and engagement with content, which are essential components of modern educational practice. Scaling digital assessments across intermediate colleges in Pakistan may improve student familiarity with technology and support self-regulated learning, provided institutions invest in teacher training, digital literacy programs, and reliable infrastructure.

Future research should explore long-term effects, adaptive digital

quizzes, and larger sample sizes to better understand the conditions under which digital assessments may produce measurable learning gains.

Recommendations and Future Implementation

The findings of this study suggest that digital self-assessment quizzes are associated with higher overall student performance and can enhance engagement and learner autonomy. To strengthen the effectiveness and sustainability of digital assessments in the Intermediate College Section, the following recommendations are proposed:

1. Educators should attend workshops and seminars on digital tools such as MS Forms. Training should focus on best practices in quiz design, interactivity, and formative assessment strategies to maximize student engagement (Graham, Borup, & Smith, 2012).
2. Implementing a system to gather students' feedback on digital assessments can help identify usability issues, technical challenges, and the perceived value of instant feedback (Nicol & Macfarlane-Dick, 2006).
3. Encouraging educators to engage in regular discussions, forums, or digital communities can promote sharing of effective strategies, co-development of assessments, and continuous improvement (Pryor & Crossouard, 2008).
4. Orientation sessions and workshops should be provided to improve students' confidence with online platforms, particularly in contexts where access to technology is limited.

By combining professional development, structured feedback, collaboration, and student support, institutions can implement digital assessments in a sustainable, pedagogically sound, and contextually appropriate manner, building a more interactive and future-ready learning environment.

References

- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–75.
- García-Peñalvo, F. J., Corell, A., Abella-García, V., & Grande, M. (2020). Online assessment in higher education in the time of COVID-19. *Education in the Knowledge Society (EKS)*, 21, Article 12. <https://doi.org/10.14201/eks.23086>
- Godsk, M., & Møller, K. L. (2024). Engaging students in higher education with educational technology. *Education and Information Technologies*, 30(6), 2941–2976. <https://doi.org/10.1007/s10639-024-12901-x>
- Graham, C. R., Borup, J., & Smith, N. B. (2012). Using TPACK as a framework to understand teacher candidates' technology integration decisions. *Journal of Computer Assisted Learning*, 28(6), 530–546. <https://doi.org/10.1111/j.1365-2729.2011.00472.x>
- Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, 54, 170–179. <https://doi.org/10.1016/j.chb.2015.07.045>
- Huang, W., Stephens, J. M., & Brown, G. T. L. (2025). Feedback assisted by technology: A systematic review of empirical research. *International Journal of Technology in Education (IJTE)*, 8(2), 421–444. <https://doi.org/10.46328/ijte.1061>
- Jamil, S., & Muschert, G. (2023). The COVID-19 pandemic and E-learning: The digital divide and educational crises in Pakistan's universities. *American Behavioral Scientist*, 68(9), 1161–1179. <https://doi.org/10.1177/00027642231156779>
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- Nikou, S., & Aavakare, M. (2021). An assessment of the interplay between literacy and digital technology in higher education. *Education and Information Technologies*, 26, 3893–3915. <https://doi.org/10.1007/s10639-021-10451-0>

- Pryor, J., & Crossouard, B. (2008). A socio-cultural theorisation of formative assessment. *Oxford Review of Education*, 34(1), 1–20. <https://doi.org/10.1080/03054980701476386>
- Saleem, F., Chikhaoui, E., & Malik, M. I. (2024). Technostress in students and quality of online learning: Role of instructor and university support. *Frontiers in Education*, 9, Article 1309642. <https://doi.org/10.3389/feduc.2024.1309642>
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78 (1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Simon, P. D., & Zeng, L. M. (2024). Behind the scenes of adaptive learning: A scoping review of teachers' perspectives on the use of adaptive learning technologies. *Education Sciences*, 14 (12), 1413. <https://doi.org/10.3390/educsci14121413>
- Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, 68, 1-16. <https://doi.org/10.1007/s11423-019-09701-3>
- Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners* (2nd ed.). ASCD.
- Waqar, Y., Rashid, S., Anis, F., & Muhammad, Y. (2024). Digital divide & inclusive education: Examining how unequal access to technology affects educational inclusivity in urban versus rural Pakistan. *Journal of Social & Organizational Matters*, 3(3), 1-13. <https://doi.org/10.56976/jsom.v3i3.97>
- Zhang, D., Zhao, J. L., Zhou, L., & Nunamaker, J. F. Jr. (2004). Can e-learning replace classroom learning? *Communications of the ACM*, 47(5), 75–79. <https://doi.org/10.1145/986213.986216>
- Zheng, M., Bender, D., & Lyon, C. (2021). Online learning during COVID-19 produced equivalent or better student course performance as compared with pre-pandemic: empirical evidence from a school-wide comparative study. *BMC Medical Education*, 21(1), Article 495. <https://doi.org/10.1186/s12909-021-02909-z>