

IN SEARCH OF STUDENTS' PREPARATION STRATEGIES: THE CONTRIBUTION OF ONLINE QUIZZES AND GRADED ASSIGNMENTS TO EXAM SUCCESS

A. Kirsh, M. Sulamy, I. Gaber

The Academic College of Tel-Aviv-Yaffo (ISRAEL)

Abstract

This study (N=113) examined predictors of exam performance in an undergraduate Operating Systems (OS) course for computer science majors. The course included graded homework and home-based, repeatable online quizzes designed to assess students' practical understanding of key OS concepts: processes, memory management, file systems, and concurrency. We compared several potential predictors of exam grades: homework grades, online quiz scores, time spent on quizzes, number of quiz attempts, and students' self-reported extent to which the quizzes contributed to their confidence in material comprehension. Quiz score emerged as the strongest individual predictor of exam performance ($r=0.311$, $R^2=9.7\%$, $p<0.005$), outperforming homework grade ($r=0.191$, $R^2=3.7\%$, $p=0.058$). When combined, all predictors explained only 14.7% of the variance in exam results. Students' quiz-attributed confidence showed no meaningful predictive value. Implications for the role of online home quizzes are discussed, addressing potential reasons for the modest correlations, and suggesting directions for future research on the role and contribution of online quizzes as formative assessment tools in supporting students' materials comprehension and exam preparation.

Keywords: Education, Computer Science, Online Quizzes, Assessment, Practice.

1 INTRODUCTION

Teaching an undergraduate Operating Systems (OS) course for computer science majors presents distinct pedagogical challenges, primarily because students often struggle with abstract and complex topics that require a deep conceptual understanding beyond basic memorization. To address these difficulties and encourage active study beyond the course homework, we developed and integrated home-based online quizzes designed to measure students' practical OS knowledge and their grasp of core materials learned in class, specifically focusing on the course topics of processes, memory management, file systems, and concurrency. Administered through the Moodle platform, these quizzes were configured to allow repeatable submissions for each student. Beyond providing a practice platform for the students, the digital nature of the quizzes enabled us to systematically collect engagement data, which assisted the teaching staff in tracking progress and following up on the overall class pace. Throughout the semester, we observed students actively utilizing the tool, frequently repeating the quizzes and investing significant time into them. These observations led to our primary research question: what is the true effect of these online quizzes on students' ultimate exam success? To investigate this, we conducted a regression analysis based on the continuously collected data, specifically examining predictors such as online quiz scores, total time spent on the quizzes, and the number of quiz attempts. Alongside this quantitative data, we also conducted a survey asking students to report on the quizzes' contribution to their learning, confidence, and overall material comprehension. The results of both the data analysis and the subjective survey are discussed herein to evaluate the effectiveness of such tools.

The integration of continuous assessment as a mechanism for enhancing learning has received significant attention in various disciplines [1], and specifically in computer science education [2]. In contrast to summative assessments, formative tools such as the online home-based quizzes utilized in this study, allow students to identify knowledge gaps in real time and receive immediate feedback, fostering active engagement with course material.

Research on retrieval practice demonstrates that active recall improves long-term retention compared to passive study [3]. This is particularly relevant in computer systems education, where students often

struggle with abstract and complex topics such as concurrency and memory management, which require deep conceptual understanding beyond rote reading [4].

Cornerstone of modern assessment theory is the “Testing Effect,” which posits that the act of retrieving information from memory significantly strengthens long-term learning compared to passive restudying. Studies by Roediger and Karpicke [3] demonstrate that self-testing acts as a powerful encoding mechanism. However, the efficacy of this effect is often mediated by the cognitive effort involved; when quizzes are perceived as repetitive or non-challenging, a state of “diminishing returns” may occur. Providing students with repeatable online quizzes is a common strategy to encourage self-regulated learning (SRL). By allowing multiple submissions, as implemented in the Moodle-based OS quizzes, instructors grant students the autonomy to correct errors and achieve mastery at their own pace. Nicol [5] argues that this iterative process increases motivation and internal feedback loops.

Recent literature suggests that immediate feedback in formative multiple-choice quizzes improves performance on similar tasks and supports near transfer of knowledge [6]. However, the extent to which such feedback supports deeper conceptual learning remains dependent on the type and richness of feedback provided.

2 DATASET AND METHODOLOGY

The dataset comprises comprehensive records from the Spring 2026 Operating Systems course at Tel Aviv-Yaffo Academic College, collected between October 2025 and January 2026. It includes 121 students with 119 exam grades ($M=49.6$, $SD=15.8$, grades range between 6-85), homework (HW) grades ($M=93.9$, $n=119$), survey responses ($n=117$) capturing students’ subjective attitudes toward quizzes, and data from three online quizzes totaling 767 attempts across 107 unique students. Quiz metrics included, for each student per quiz: maximum score achieved (across all submissions for this quiz), total time invested, and attempt counts. Raw quiz data required cleanup due to implausible short and long durations. Entries under 2 minutes were removed as insufficient for meaningful engagement, preventing artificial inflation of attempt counts. For excessively long sessions likely reflecting idle time, we capped durations at a “reasonable completion threshold” calculated as 10 minutes per answered question (regardless of correctness) + one minute per unanswered question, following established educational data analysis practices [7].

This time adjustment affected 115 entries (14.97%), with an average cap of 49.4 minutes. The high capping rate reflects prevalent long idle sessions; capping preserves these data points while preventing time invested inflation. Additionally, entries exceeding 50 minutes with zero or single-question submissions were excluded as unreliable. Post-cleaning yielded 103 students and 652 quiz attempts. For integrated analysis combining quiz, survey, HW, and exam data, we focused on the intersection of students with complete records across all sources: 99 students with quiz participation (17 took 1 quiz, 42 took 2, 40 took all 3) and 605 attempts; +14 quiz avoiders with HW and exam data.

Table 1 summarizes data for analyzed students who completed at least one quiz. Table 2 summarizes data for analyzed students who completed 0 quizzes.

Table 1. Descriptive statistics for Quiz Takers (N=99)

Measure	Exam Grade	HW Grades	Quiz Time	Quiz Attempts	Quiz Scores
Mean	50.1	95.3	289	6.07	65.9
Median	50	97	197	4	75
Standard deviation	15.1	6.28	282	6.2	23.3
Minimum	6	67	9.87	1	0
Maximum	85	100	1353	38	96.7

Table 2. Descriptive statistics for Quiz Avoiders (N=14)

Measure	Exam Grade	HW Grades
Mean	50.5	88.7
Median	49	94
Standard deviation	16.2	12.3
Minimum	19	66
Maximum	82	99

3 ANALYSIS

Home-repeatable online quizzes serve as formative assessment tools, enabling students to practice practical OS concepts (processes, memory management, concurrency) at their own pace while receiving immediate feedback. We hypothesize that greater quiz engagement, measured through participation, time invested, and performance, correlates with exam success. The rationale is that quiz work enhances exam readiness through repeated exposure, strengthening problem-solving transfer and conceptual retention for improved summative outcomes.

This chapter presents three complementary analyses. Section 3.1 compares exam performance by number of quizzes taken (0, 1, 2, or 3), contrasting quiz avoiders with varying participation levels. In Section 3.2 we examine whether self-reported quiz confidence (via survey) predicts exam success. Then, Section 3.3 employs linear regressions to compare HW grades vs. quiz metrics (scores, time, attempts) as predictors and quantify their separate and combined explanatory power.

Processing this data will reveal quizzes' impact on students' success: whether they merit retention, require improvement, or demonstrate value in repeatability (attempt contributions). This analysis may help in clarifying quizzes' formative value and guide optimal implementation strategies for technical courses.

3.1 Num Quizzes Performed

This section examines whether quiz participation breadth correlates with exam success. We categorize students by number of distinct quizzes attempted (0, 1, 2, or 3), comparing quiz avoiders (n=14) with those showing partial engagement (1 quiz: n=17; 2 quizzes: n=42) versus full engagement (3 quizzes: n=40). We analyze exam grades and quiz engagement patterns (average scores, time invested, number of attempts) across these groups to assess whether broader quiz exposure yields measurable exam performance gains. Our hypothesis is simple: students practicing more quizzes during the semester (all 3 vs. 2; 2 vs. 1; 1 vs. 0) should show progressively higher exam grades.

Table 3 compares the groups. To our surprise, exam grades show no clear difference across them. The average exam grade is very similar in all four groups, with a slightly higher, though non-significant, mean for students who took all three quizzes (51.3, versus 49.0 for two quizzes, 50.2 for one quiz, and 50.1 for no quizzes).

Table 3: Results by number of quizzes performed (N=113)

Quizzes Performed	N	Exam Grade	HW Grades	Quiz Time	Quiz Attempts	Quiz Scores
0	14	50.1	95.3	–	–	–
1	17	50.2	96.8	101.2	1.9	58.0
2	42	49.0	95.4	183.6	3.9	62.6
3	40	51.3	94.5	480.0	10.2	72.8

Figure 1 presents the distribution of exam grades across the groups, visually showing that the differences between the groups are minimal.

At first glance, this seems counterintuitive, as if taking the quizzes had little effect at all. However, the result suggests that merely entering a quiz is not enough for success; what matters is actual engagement, reflected in the time and effort invested in each quiz. Students who spent little time or made minimal effort would not necessarily improve their exam performance. In many cases, students may have opened a quiz only to view the questions, and such cases may still count as attempts if they spent at least 2 minutes in the quiz. A more detailed analysis of actual quiz engagement and impact is therefore needed to understand the quizzes' value and contribution, as we will do in the following sections.

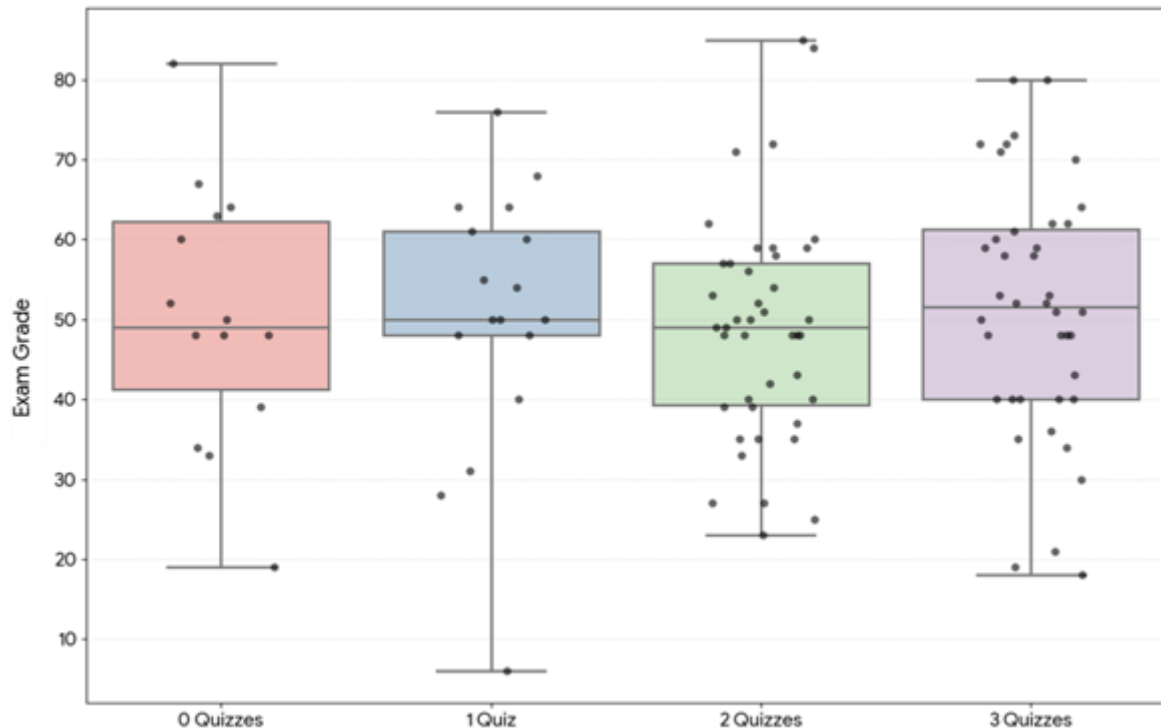


Figure 1: Exam Grade Distribution by Number of Quizzes Taken

3.2 Self-reported quiz confidence

Prior work in computer science education has examined the relationship between self-reported confidence in quiz performance and actual learning outcomes. Lee and Liao [8] investigated confidence estimates attached to self-assessment quizzes and found that, although repeated quiz practice correlated with improved exam performance, students' confidence reports did not necessarily improve metacognitive calibration or accurately reflect actual understanding.

In this section, we analyze the relationship between reported confidence following quiz work and exam success. This analysis helps us better understand the extent to which students' subjective assessment of their performance aligns with their actual results, or, on the other hand, whether quiz work may boost confidence without necessarily reflecting true mastery.

In addition to measuring students' quiz performance and statistics, we also examined their self-reported confidence, as reported following their work on the quiz questions. This was done through a retrospective question in a survey conducted after the final exam, before students received their exam results. We asked the students: "How did the quizzes affect your level of confidence in understanding the material?" with the following possible answers: (a) my confidence in understanding the material increased; (b) my confidence in understanding the material decreased; (c) my confidence didn't change, it was good and remained; (d) my confidence didn't change, it was poor and remained; (e) I didn't perform quizzes.

Confidence ratings provide an important complementary perspective, as they reflect how students perceive the impact of the quizzes on their understanding. We tend to assume that reporting improved confidence would indicate a positive effect of the quiz and may be associated with higher exam grades, whereas lower confidence may suggest that the quiz did not support learning effectively and may be associated with lower exam grades.

One might argue that if quizzes lower confidence, students would invest more time in learning; however, since the survey was conducted at the end of the semester, the retrospective effect would more likely reflect how students felt after repeating the quizzes. If the quizzes helped them grasp the material, this would be reported as improved confidence and would be expected to correlate positively with exam results.

Based on the above, our hypothesis was that students reporting improved confidence following their quiz work would, on average, have higher exam grades than those reporting a loss of confidence. The results are presented in Table 4. As can be seen, self-reported confidence is not a good predictor of exam success. One group stands out above the others: the group marking “my confidence didn’t change, it was good and remained.” However, a quick look at the data reveals that this group consists of only one student. We conclude that, at least in this analysis, student self-reported confidence based on quiz work cannot help predict exam success. This should be taken into account by instructors, as we sometimes rely on asking students, “Are you okay? Do you feel ready for the exam?” and, as it seems, relying on their answers may not be a good predictor. This suggests that better predictors may be the actual score or students’ engagement in the quizzes. Which is our next step in the following section.

Table 4: Survey analysis by changes in confidence levels (N=99)

Confidence Category	N	Exam Grade	HW Grades	Num. Quizzes	Quiz Time	Quiz Att.	Quiz Score
Confidence Increased	30	51.5	95.4	2.3	349.6	7.7	72.3
Confidence Decreased	61	50.4	94.9	2.2	288.1	5.7	65.0
Unchanged - High	1	59.0	98.5	3.0	207.4	5.0	76.7
Unchanged - Low	3	47.3	97.3	2.0	67.5	3.3	26.1
Others*	4	35.0	96.9	1.3	40.2	1.5	58.8

* Students who reported no quiz activity despite actually performing them.

3.3 Regression Model

In this section, we examine the relationship between course performance measures and exam grades using regression analysis. In addition to quiz-related variables, we also include homework grades (HW) in order to assess their association and strength with exam performance.

We first conducted a single-variable correlation analysis. The results, shown in Table 5, indicate that exam grade is positively correlated with all the predictors, but the relationship is stronger for the quiz variables than for homework grades. Quiz scores show the strongest association with exam grade ($r = 0.311$, $p = 0.002$), followed by quiz attempts ($r = 0.255$, $p = 0.011$) and quiz time ($r = 0.243$, $p = 0.015$), while homework grades show only a weaker and marginal relationship ($r = 0.191$, $p = 0.058$). In variance terms, quiz scores alone explain about 9.7% of the variance in exam grades, compared with about 6.5% for quiz attempts, 5.9% for quiz time, and only 3.6% for homework grades. This suggests that quiz-related measures have greater predictive power than homework grades when considered separately.

Table 5: Correlation matrix of exam performance and study variables (N=99)

Variable	Statistic	Exam Grade	HW Grades	Quiz Scores	Quiz Time
HW Grades	Pearson's r	0.191	–		
	p-value	0.058	–		
Quiz Scores	Pearson's r	0.311**	0.127	–	
	p-value	0.002	0.210	–	
Quiz Time	Pearson's r	0.243*	-0.088	0.475***	–
	p-value	0.015	0.388	<0.001	–
Quiz Attempts	Pearson's r	0.255*	-0.048	0.388***	0.878***
	p-value	0.011	0.636	<0.001	<0.001

Note. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The single-variable analysis also reveals strong correlations among the quiz variables themselves. For example, quiz time and quiz attempts are very strongly correlated ($r = 0.878$, $p < 0.001$), and quiz scores are also strongly related to quiz time ($r = 0.475$, $p < 0.001$) and quiz attempts ($r = 0.388$, $p < 0.001$). This indicates substantial overlap between the quiz measures, meaning that they do not provide fully independent information. Therefore, in the multiple linear regression model, we included only the variables that add relatively independent information, in order to reduce multicollinearity and estimate each variable's unique contribution more clearly.

We then conducted a multiple linear regression model using homework grades, quiz scores, and quiz attempts. The model was statistically significant overall ($F = 5.44$, $p = 0.002$), but its explanatory power remained limited ($R^2 = 0.147$). This means that the model explains 14.7% of the variance in exam grades, while 85.3% remains unexplained. Within the model, quiz scores emerged as the only statistically significant predictor of exam grade ($B = 0.14$, $p = 0.037$), while homework grades ($B = 0.41$, $p = 0.078$) and quiz attempts ($B = 0.43$, $p = 0.089$) showed positive but non-significant trends. Standard OLS (Ordinary Least Squares) assumptions were tested and met: we checked that the residuals are normally distributed (Shapiro–Wilk $p = 0.864$, indicating that they are). Tests for heteroskedasticity were non-significant (e.g., Breusch–Pagan $p = 0.546$), suggesting homoscedasticity. Multicollinearity was not a concern, as variance inflation factors (VIFs) were low (all VIFs < 1.25).

In terms of predictive contribution, the results again suggest that quiz performance carries more explanatory value than homework grades. The results of the multiple linear regression model are detailed in Table 6.

Figure 2 presents the comparison of predictive power across the variables and models. As shown, the quiz measures outperform homework grades in predicting exam success, and this pattern remains visible in the multiple regression model as well. At the same time, the large share of unexplained variance suggests that other factors are also shaping exam performance. These may include prior knowledge, study habits, attendance, motivation, exam anxiety, and other personal or contextual differences. Among these, some are outside the instructor's control, but others may be influenced by course design, such as the structure and timing of quizzes, the quality of homework feedback, and the amount of practice and support provided during the semester. Overall, the results suggest that quiz performance is a more informative indicator of exam achievement than homework grades, although neither set of measures can fully explain exam outcomes.

Table 6: Multiple linear regression analysis for predicting exam grades

Predictor	B	Std. Error	β	t	p
Intercept	-1.16	21.88	–	–	–
HW Avg	0.41	0.23	0.17	1.78	0.078
Quiz Scores	0.14	0.07	0.22	2.12	0.037*
Quiz Attempts	0.43	0.25	0.18	1.72	0.089

*B = unstandardized coefficient; β = standardized coefficient.
The overall model was significant, $F = 5.44$, $p = 0.002$, $R^2 = 0.147$. * $p < 0.05$*

4 DISCUSSION

This study analyzes data from an Operating Systems course's home-repeatable online quizzes (after dataset cleanup: number of quiz attempts = 605, quiz takers = 99, quiz avoiders = 14), alongside homework grades, exam scores and survey responses on students' confidence following quiz work, to identify predictors of final exam performance.

While online quizzes demonstrate a statistically significant positive effect on exam performance, their explanatory power is modest, far less than expected for such intensive engagement (both for course instructors as well as for the students). The combined model (quiz score, time, homework) explains only 14.7% of exam variance. These findings challenge the assumption that graded practice volume and performance directly translate to exam mastery in technical courses like Operating Systems. It indicates that core drivers of exam success, likely deep conceptual understanding from lectures, prior programming proficiency, problem-solving transfer skills, and exam-specific preparation strategies operate independently of measurable quiz engagement.

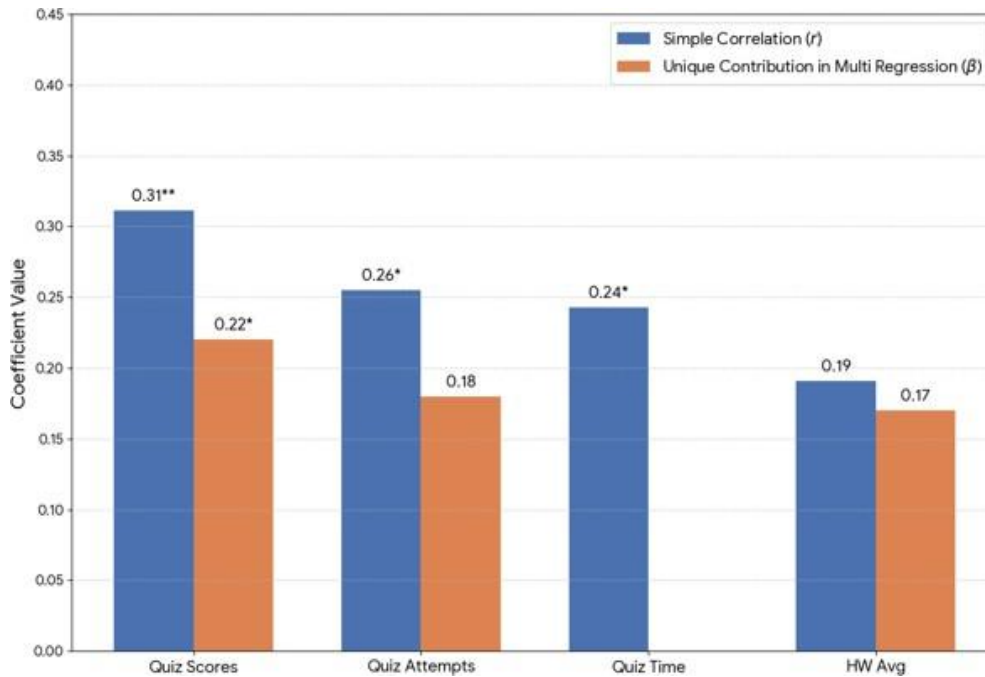


Figure 2: Comparison of simple correlations and standardized regression weights
Significance: * $p < 0.05$, ** $p < 0.01$.

Notes: Bars compare standalone relationship strength (Simple Correlation, r) with unique predictive weight (Standardized β) once other factors are controlled. The overall regression model explains 14.7% of exam grade variance. ($R^2 = 0.147$, $p = 0.002$).

Quiz Time was excluded from the regression analysis to prevent multicollinearity due to its high correlation with Quiz Attempts ($r = 0.88$).

The modest predictive power of home-based quiz performance on final exam success ($R^2 = 9.7\%$) suggests that simply engaging with the quizzes does not guarantee mastery. These results can be elucidated through the framework of Bjork and Bjork [9], who distinguish between “performance improvement”, as the observable improvement during practice, and “learning”, as the long-term, stable acquisition of knowledge. In our context, multiple quiz attempts may represent an improvement in immediate performance due to item familiarity, rather than a genuine gain in stable learning that can be transferred to the exam environment.

The study shows the importance of quizzes as predictors of exam results, though their explanatory power is modest. The results suggest limited transfer from practice to assessment. We therefore recommend better aligning quiz formats with exam style, analyzing quiz failures during the semester to proactively support students, and practicing in-class quizzes in an exam-like environment (limited time, one attempt, close materials), as better preparation for the exam. Further research may focus on the contribution of home-quizzes to success on in-semester exam-like practice tests, seeking to better explain exam performance through lecture engagement patterns, problem-solving transfer skills, and quizzes as complementary predictors. This approach would help isolate quiz impact from self-exam preparation, clarifying the actual effect of online quizzes on knowledge acquisition.

REFERENCES

- [1] P. Black and D. Wiliam, *Inside the black box: Raising standards through classroom assessment*. Granada Learning, 1998.
- [2] P. Ihantola, T. Ahoniemi, V. Karavirta, and O. Seppälä, "Review of recent systems for automatic assessment of programming assignments," in *Proceedings of the 10th Koli Calling International Conference on Computing Education Research*, pp. 86–93, 2010.
- [3] H. L. Roediger and J. D. Karpicke, "Test-enhanced learning: Taking memory tests improves long-term retention," *Psychological Science*, vol. 17, no. 3, pp. 249–255, 2006.
- [4] S. Pamplona, N. Medinilla, and P. Flores, "Exploring misconceptions of operating systems in an online course," in *Proceedings of the 13th Koli Calling International Conference on Computing Education Research*, pp. 77–86, 2013.
- [5] D. J. Nicol and D. Macfarlane-Dick, "Formative assessment and self-regulated learning: A model and seven principles of good feedback practice," *Studies in Higher Education*, vol. 31, no. 2, pp. 199-218, 2006.
- [6] A. Ryan, T. Judd, D. Swanson, D. P. Larsen, S. Elliott, K. Tzanetos, and K. Kulasegaram, "Beyond right or wrong: More effective feedback for formative multiple-choice tests," *Perspectives on Medical Education*, vol. 9, no. 5, pp. 307–313, 2020.
- [7] V. Kovanovic, D. Gašević, S. Dawson, S. Joksimovic, and R. Baker, "Does time-on-task estimation matter? Implications on validity of learning analytics findings," *Journal of Learning Analytics*, vol. 2, no. 3, pp. 81-110, 2015.
- [8] P. Lee and S. N. Liao, "Targeting metacognition by incorporating student-reported confidence estimates on self-assessment quizzes," in *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, pp. 431-437, 2021.
- [9] R. A. Bjork and E. L. Bjork, "A new theory of disuse and an old theory of stimulus fluctuation," in *From learning processes to cognitive processes: Essays in honor of William K. Estes*, vol. 2, pp. 35–67, 1992.