Research Article

Unannounced Quizzes and Exam Performance: Evidence of a Continuous Preparation Effect in Introductory Microeconomics

Moon Joon Kim[a]
[a]George Mason University Korea

JEL Codes: A22, I23, D01
Keywords: best-of-*N* grading, preparation effect, satisficing, strategic slacking, unannounced quizzes

Abstract
This study investigates whether unannounced quizzes can improve student learning outcomes in introductory microeconomics courses. Using data from 257 students and controlling for section-by-semester fixed effects, we find that unannounced quizzes significantly improve final exam performance. Moreover, quiz performance predicts exam success even when quiz scores do not count toward the final grade. This suggests that the results are driven by a continuous preparation effect—students maintaining regular study habits in anticipation of surprise assessments—rather than by immediate grade incentives. However, under a "best-of-*N*" quiz grading policy, we identify a trade-off. While this policy provides a safety net for lower-performing students (consistent with mastery learning principles), it inadvertently encourages strategic slacking among high achievers. Once these students have secured the maximum possible quiz credit, they rationally reduce their effort. These findings underscore the importance of designing assessment structures that sustain motivation across the full spectrum of student ability levels.

1 Introduction

The role of quizzes in educational settings has been explored extensively, with a growing consensus that frequent assessment positively influences student performance. Much of this literature is grounded in the "testing effect," which posits that the act of retrieving information strengthens memory cues and enhances long-term retention compared to passive review techniques (Roediger and Karpicke 2006; Carpenter et al. 2008; Karpicke and Roediger 2008). While the cognitive benefits of retrieval are well-documented, the behavioral mechanisms driven by the anticipation of assessments—specifically unannounced assessments (hereafter referred to as pop quizzes)—remain less understood. Unlike scheduled exams or assignments that often induce deadline-driven study habits or massed practice (i.e., cramming), unannounced quizzes introduce an element of unpredictability. This study leverages this feature to examine the "preparation effect," investigating whether the uncertainty of pop quizzes encourages students to maintain consistent engagement with course material throughout the semester (Kamuche 2007; Blasco-Arcas et al. 2013).

The distinction between announced and unannounced assessments is critical in pedagogical design. Previous research suggests that announced quizzes, while beneficial, may still allow for procrastination. Conversely, unannounced quizzes are theorized to foster a heightened sense of urgency and focus, potentially driving more consistent study habits typically absent in scheduled assessments (Kamuche 2007; Azorlosa 2011). However, empirical evidence on their efficacy is mixed. While some studies contend that pop quizzes are integral to developing effective learning and metacognitive strategies (Karpicke et al. 2009; Azorlosa 2011), others argue that they may be counterproductive. Critics suggest that the element of surprise can heighten student anxiety and diminish intrinsic motivation, particularly for those with lower test-taking confidence, leading to outcomes that reflect stress rather than knowledge (Putwain and Symes 2011; Segool et al. 2013; Stowell 2015). Furthermore, effective implementation often depends on the nature of the subject matter and integration into the overall teaching strategy (Bangert-Drowns et al. 1991).

In the context of economics education, where cumulative knowledge building is essential, validating the impact of such assessments is particularly important. Recent scholarship has examined various assessment tools; for instance, Latif and Miles (2020) found that while graded homework significantly improved academic outcomes in economics courses, the specific role of unannounced quizzes requires further empirical scrutiny. This study seeks to fill this gap by quantifying the effects of unannounced quizzes on final exam performance in introductory microeconomics courses. We aim to determine whether the incentives provided by surprise assessments translate into improved academic achievement, providing insight into the ongoing pedagogical debate (Black and Wiliam 1998; Wiliam 2011).

To address this question, we analyze a comprehensive dataset of 257 students enrolled in introductory microeconomics classes over four semesters. A key feature of our study design is the implementation of a specific grading policy intended to mitigate the potential negative effects of anxiety: While students faced ten unannounced quizzes, only the top two scores contributed to their final grade. This "best-of-*N*" approach aligns with the principles of mastery learning, designed to balance the rigor of continuous assessment with the flexibility needed to reduce high-stakes stress (Bloom 1968; Guskey 2007). Utilizing section-by-semester fixed effects to control for temporal shocks and instructor variability (Andrews 2005; Wooldridge 2015), our empirical strategy isolates the impact of quiz performance from other confounding factors.

Our findings provide robust evidence supporting the pedagogical utility of unannounced quizzes. We find that higher performance in unannounced quizzes is significantly associated with improved final exam scores, suggesting that the anticipation of surprise assessments indeed promotes consistent preparation and deeper engagement. These results are robust to a series of sensitivity analyses, including alternative specifications and subsample tests. By highlighting the preparation effect in an economics context, this study underscores the benefits of applying cognitive psychology principles to foster proactive learning while managing student anxiety through thoughtful course design (Bjork et al. 2013).

2 Background

This section provides the theoretical and contextual foundation for our study. We begin by outlining the conceptual framework rooted in cognitive psychology, specifically focusing on the mechanisms of the preparation effect and desirable difficulties. We then describe the specific course structure and the "best-of-*N*" assessment design employed in the introductory microeconomics course to clarify the empirical setting.

2.1 Conceptual Framework: The Preparation Effect and Desirable Difficulties

Central to our study is the concept of the preparation effect, a principle rooted in cognitive psychology that plays a pivotal role in educational practices. Unlike the testing effect, which emphasizes the cognitive benefits of retrieval practice for long-term memory retention (Roediger and Karpicke 2006; Karpicke and Roediger 2008), the preparation effect focuses on the motivational and behavioral changes induced by the anticipation of assessments (Kamuche 2007). This effect suggests that the possibility of facing unexpected quizzes encourages students to maintain a continuous and proactive engagement with their studies, fostering a consistent study habit that leads to improved academic performance (Mawhinney et al. 1971; Graham 1999).

The mechanism behind the preparation effect can be conceptually situated within theories of self-regulated learning and metacognition. These frameworks emphasize how the anticipation of evaluation shapes proactive study behaviors, encouraging students to monitor their learning progress continuously (Zimmerman 2002; Bjork et al. 2013).

Theoretical support for this approach is further strengthened by the concept of "desirable difficulties" (Bjork 1994; Bjork and Bjork 2011). This framework posits that learning tasks requiring considerable cognitive effort—such as responding to unexpected queries—often lead to more durable learning than easier, predictable tasks.

While assignments and scheduled exams motivate preparation, their predictable timing often allows students to concentrate effort in short, intensive periods (i.e., cramming) rather than sustaining engagement (Cepeda et al. 2006). By contrast, pop quizzes introduce uncertainty and immediacy; because students cannot anticipate when they will occur, they are encouraged to study continuously to remain prepared at all times (Leeming 2002).

However, the implementation of such uncertainty requires careful consideration of student anxiety. The Yerkes–Dodson Law (Yerkes and Dodson 1908) suggests an inverted U-shaped relationship between arousal and performance; while moderate arousal can enhance focus, excessive anxiety acts as a cognitive load that impairs performance (Zeidner 1998; Cassady and Johnson 2002). Therefore, the pedagogical challenge lies in leveraging the preparation effect to induce distributed study habits while mitigating the debilitating anxiety that can hinder learning (Putwain 2008).

2.2 Key Features of the Course Syllabus and Assessment Design

In the realm of economics education, assessment methodologies play a pivotal role in shaping student engagement. The introductory microeconomics course analyzed in this study employs a multifaceted evaluation schema: a midterm exam (20%), a final exam (30%), writing assignments (20%), attendance (10%), and quizzes (20%). This structure aims to holistically assess a student's grasp of the course material through various pedagogical lenses (Walstad and Becker 1994).

The central component of this study is the use of quizzes in an unannounced format. This approach differs fundamentally from traditional top-down assessments, such as scheduled homework or group projects, where instructors set fixed timing for evaluation. Such methods often position students as passive recipients who comply with externally imposed deadlines (Boud 2000).

In contrast, unannounced quizzes represent a bottom-up approach, emphasizing continuous engagement. By introducing short, frequent, and unpredictable assessments, the course encourages students to adopt regular study habits in anticipation of possible evaluations, thereby fostering intrinsic motivation and self-regulation (Black and Wiliam 1998).

A critical aspect of this assessment method is the best-of-*N* grading policy. Throughout the semester, students are exposed to ten unannounced quizzes, but only the top two quiz grades are factored into the final course grade. This policy is grounded in the principles of mastery learning (Bloom 1968; Guskey 2007), which prioritizes the attainment of competence over the timing of learning.

By allowing students to drop their lowest scores, the course treats early failures as formative opportunities rather than summative penalties. This strategy ensures a level of unpredictability to drive the preparation effect while providing a safety net to reduce high-stakes anxiety. Consequently, students are incentivized to maintain consistent attendance and engagement without the paralyzing fear of a single poor performance affecting their overall grade.

3 Methods

This section outlines the research methodology employed to examine the impact of unannounced quizzes on student performance. We first describe the data collection process and the operationalization of key variables, including the specific assessment structure. Subsequently, we detail the empirical strategy, which utilizes a fixed-effects regression framework to ensure robust causal inference.

3.1 Data

This study utilizes a comprehensive dataset from an introductory microeconomics course offered at a branch campus of a US university in South Korea, spanning four semesters from Spring 2022 to Fall 2023. The data collection features repeated cross-sectional observations. Each semester included two identical class sessions, each 75 minutes in duration, ensuring consistency in instructional delivery.

The initial enrollment records included 277 students. However, to ensure consistency across all descriptive statistics and regression analyses, we restrict our final analytical sample to 257 students for whom complete information on all covariates—including midterm scores, attendance, assignments, and demographic variables—is available. Specifically, we excluded five students who did not submit assignments, two students who missed the midterm exam, and 13 students who were absent for the final exam. This uniform sample size (N = 257) allows for direct comparability across all tables and models presented in this study.

Table 1 provides descriptive statistics for this final sample. A key focus is on the midterm and final exam scores, attendance records, assignment grades, and pop quiz performance. The pop quiz score represents the average score across all ten unannounced quizzes administered during the semester. While only the top two quiz grades counted toward the official course grade, we use the average score in our analysis to capture the cumulative level of student engagement and preparation.

Table 1. Descriptive statistics

	N	Mean	SD	Min.	Max.
Midterm Exam	257	72.09	15.92	30	100
Final Exam	257	72.93	16.63	27	100
Attendance	257	8.33	2.44	0	10
Assignments	257	8.77	0.97	4	10
Pop Quiz	257	2.92	0.68	1.40	4.78
Female	257	0.47	0.50	0	1
Transfer	257	0.18	0.39	0	1

Notes: *Midterm exam* and *Final exam* are measured on a 100-point scale. *Attendance* is measured on a 0–10 scale, with 1 point deducted for each absence. *Assignments* represents the average score of three semester tasks (0–10 scale). *Pop Quiz* denotes the average score across ten unannounced quizzes (0–5 scale). *Female* is a binary variable equal to 1 if the student is female, and 0 otherwise. *Transfer* takes the value of 1 if the student transferred from another institution.

Attendance serves as a proxy for student engagement and exposure to instructional material. The average attendance score is 8.33 out of 10 (SD = 2.44), with 1 point deducted for each unexcused absence. Consistent with the literature suggesting a strong link between class attendance and academic performance (Romer 1993; Stanca 2006), this variable controls for the direct effect of instructional time on exam outcomes.

Assignments played a formative role in the course assessment. Three writing assignments were issued throughout the semester, designed to reinforce lecture material by requiring students to apply economic concepts to real-world examples, such as analyzing news articles or explaining incentives in case studies (Walstad 2001). The average assignment score is 8.77 out of 10 (SD = 0.97), indicating a high level of completion and understanding among the retained sample.

It is worth noting that the data period (2022–2023) precedes the widespread integration of advanced generative AI tools in the classroom; Thus, assignment scores are considered authentic measures of student effort rather than AI-assisted outputs. These assignments contributed 20% to the final grade and functioned primarily as "assessment for learning" (Black and Wiliam 1998).

The midterm exam scores have an average of 72.09 (SD = 15.92), while the final exam scores show a slightly higher mean of 72.93 (SD = 16.63). This distribution suggests a broad dispersion in performance, indicating varying levels of student preparedness as the course progressed.

3.2 Empirical Strategy

This study aims to identify the causal impact of unannounced quizzes on student learning outcomes. While randomized controlled trials (RCTs) are the gold standard for causal inference, educational settings often require quasi-experimental approaches due to ethical and logistical constraints. Following the econometric framework for repeated cross-sectional data suggested by Andrews (2005) and Wooldridge (2010), we employ an ordinary least squares (OLS) regression model with section-by-semester fixed effects.

Since the data are not derived from random assignment, it is crucial to examine potential selection bias or pre-existing differences in student ability. To address this, we first estimate a model examining the determinants of midterm exam performance, which serves as an intermediate outcome relatively less affected by the cumulative preparation effect of quizzes compared to the final exam:

$$Midterm_i = \alpha + \gamma \cdot Z_i + \eta_c + \epsilon_i \,. \tag{1}$$

In equation (1), Z_i represents a vector of student characteristics, including gender, major, transfer status, year of admission, and residential location; η_c denotes section-by-semester fixed effects. This analysis, reported in Table 2, functions as a balance check to verify whether demographic factors systematically influence academic performance (Imbens and Wooldridge 2009).

To test our primary hypothesis—that unannounced quizzes improve final exam performance—we estimate the following equation:

$$FinalExam_i = \alpha + \beta_1 \cdot PQ_i + \beta_2 \cdot X_i + \beta_3 \cdot Z_i + \eta_c + \epsilon_i \,. \tag{2}$$

Here, $FinalExam_i$ is the dependent variable. The key independent variable, PQ_i, is the average score of all ten unannounced pop quizzes. Although only the top two scores determined the course grade, we utilize the average of all ten quizzes to capture the intensity of continuous preparation throughout the semester. X_i includes time-varying academic controls such as attendance and assignment scores, while Z_i controls for time-invariant student characteristics.

We include section-by-semester fixed effects (η_c) to absorb unobserved heterogeneity specific to each class session (e.g., 9:00AM vs. 10:30AM learning environments) and semester-level shocks (e.g., variations in exam difficulty or academic calendar). This strategy isolates the effect of quizzes from broader contextual factors. Furthermore, to account for potential heteroscedasticity and serial correlation within student groups, we cluster standard errors at the student major level or residential level in our robustness checks, following the recommendations of Cameron and Miller (2015).

To ensure the stability of our estimates, we rigorously tested for multicollinearity among explanatory variables. We calculated variance inflation factors (VIFs) for all predictors. The mean VIF for our main specification is 2.40, with a maximum VIF of 4.16. These values are well below the conventional threshold of 10, confirming that multicollinearity is not a concern in our analysis (Chatterjee and Hadi 2012; Wooldridge 2015).

Finally, to provide a comprehensive assessment of model performance, we report both adjusted R^2 and root mean square error (RMSE). The adjusted R^2 penalizes model complexity, offering a more accurate measure of goodness-of-fit, while RMSE provides an estimate of the standard deviation of the prediction errors, allowing for a clearer interpretation of model precision.

This study was conducted in accordance with ethical standards for research involving human subjects. Since the analysis relies solely on de-identified retrospective educational records, the Institutional Review Board (IRB) determined that the study was exempt from formal review.

4 Results

4.1 Determinants of Midterm Grades

Table 2 presents the regression analysis of individual characteristics on midterm grades. This analysis serves a dual purpose: exploring the factors influencing initial academic performance and conducting a balance check to ensure that student outcomes are not driven by pre-existing demographic disparities. Column 1 reports the baseline OLS estimates, while columns 2 and 3 incorporate section-by-semester fixed effects and cluster-robust standard errors, respectively.

The results indicate that demographic factors, specifically gender and residential origin, do not have a statistically significant relationship with midterm performance across all model specifications. For instance, the coefficient for gender is small and statistically insignificant (β = 0.87, $p > 0.10$), suggesting a gender-neutral academic environment where neither male nor female students hold a systematic advantage. Similarly, transfer status and academic year classification do not significantly predict midterm outcomes.

However, academic major plays a substantial role. Students majoring in Economics ($\beta = 7.51$, $p < 0.01$) and Business ($\beta = 4.24$, $p < 0.01$) score significantly higher than their peers from other disciplines. This performance gap is expected and likely reflects a stronger foundational background or intrinsic interest in the subject matter, consistent with the expectancy-value theory of motivation (Eccles and Wigfield 2002).

Notably, the adjusted R^2 for these models is low (approximately 0.01), and the RMSE remains around 15.83. While a low R^2 might typically suggest limited explanatory power, in this context, it effectively serves as a validity check. It confirms that observable student characteristics—such as where they live or their gender—explain very little of the variation in exam scores. This implies that the midterm grades are largely driven by unobserved individual factors, such as study effort and aptitude, rather than by systematic demographic biases. Thus, the low explanatory power of demographic variables reinforces the fairness of the baseline assessment environment.

4.2 The Impact of Unannounced Quizzes on Final Exam Performance

Table 3 illustrates the main results regarding the effect of pop quizzes on final exam grades. Columns 1–3 present linear specifications with increasing levels of rigorous controls. Across all models, we observe a robust and statistically significant positive relationship between unannounced quiz performance and final exam scores ($p < 0.01$).

Table 2. Determinants of midterm grades

Dependent variable: Midterm grade	1	2	3
Academic background			
- Major (reference all other majors)			
Business	4.69**	4.24*	4.24***
	(2.13)	(2.17)	(1.07)
Economics	7.53*	7.51*	7.51***
	(3.98)	(4.11)	(1.48)
- Classification level (reference: freshman)			
Sophomore	4.58	3.86	3.86
	(3.09)	(3.16)	(3.43)
Junior or above	−0.62	−1.66	−1.66
	(6.01)	(6.07)	(3.80)
Transfer status	−0.81	−0.49	−0.49
	(3.17)	(3.29)	(1.94)
Demographic background			
- Sex	0.74	0.87	0.87
	(2.01)	(1.98)	(0.92)
- Permanent residential address (reference all other regions)			
Seoul	−1.65	−0.84	−0.84
	(3.30)	(3.37)	(4.29)
Incheon	−0.57	0.26	0.26
	(3.72)	(3.77)	(4.14)
Gyeonggi	2.73	3.63	3.63
	(3.44)	(3.51)	(4.87)
Foreign	−1.30	0.26	0.26
	(4.08)	(4.32)	(5.08)
Section-by-semester FE	No	Yes	Yes
Cluster-robust SE	No	No	Yes
No. of obs.	257	257	257
R^2	0.05	0.08	0.08
Adj. R^2	0.01	0.01	0.01
Root mean squared error (RMSE)	15.82	15.83	15.83

Notes: The dependent variable is the midterm exam score (0–100 scale). Column 1 presents baseline ordinary least squares (OLS) estimates. Column 2 includes section-by-semester fixed effects. Column 3 employs cluster-robust standard errors. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3. The impact of unannounced quizzes on final exam performance

	1 Linear	2 Linear	3 Linear	4 Nonlinear
Pop Quiz	10.55***	8.67***	8.67***	
	(1.31)	(1.24)	(2.22)	
Nonlinear effects (ref: $Pop\ Quiz \leq \text{Mean} - 1\text{SD}$)				
$\text{Mean} - 1\text{SD} < Pop\ Quiz \leq \text{Mean}$				6.10***
				(1.60)
$\text{Mean} < Pop\ Quiz \leq \text{Mean} + 1\text{SD}$				10.10**
				(3.53)
$\text{Mean} + 1\text{SD} < Pop\ Quiz$				16.25***
				(4.40)
Assignments	−0.17	1.58*	1.58	1.87
	(0.88)	(0.84)	(1.05)	(1.22)
Attendance	0.41	0.15	0.15	0.16
	(0.35)	(0.32)	(0.22)	(0.25)
Midterm	0.40***	0.47***	0.47***	0.51***
	(0.06)	(0.05)	(0.10)	(0.10)
Academic controls	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes
Section-by-semester FE	No	Yes	Yes	Yes
Cluster-robust SE	No	No	Yes	Yes
No. of obs.	257	257	257	257
R^2	0.54	0.65	0.65	0.63
Adj. R^2	0.51	0.62	0.62	0.60
Root mean square error (RMSE)	11.60	10.30	10.30	10.56

Notes: The dependent variable is the final exam score (0–100 scale). Columns 1–3 present linear regression estimates. Column 2 includes section-by-semester fixed effects. Column 3 employs cluster-robust standard errors. Column 4 examines nonlinear effects by categorizing quiz performance into four bins based on the mean and standard deviation, with the lowest-performing group serving as the reference. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

In the baseline model without fixed effects (column 1), the coefficient for pop quizzes is large ($\beta = 10.55$). However, after controlling for section-by-semester fixed effects (column 2) and clustering standard errors (column 3), the magnitude of the coefficient stabilizes at 8.67 ($p < 0.01$). In our preferred specification (column 3), a 1-point increase in the average quiz score (on a 5-point scale) is associated with an 8.67-point increase in the final exam score. Considering the standard deviation of the final exam (16.63), this translates to an effect size of approximately 0.52 SD, indicating a moderate-to-strong pedagogical impact.

Reviewing the goodness-of-fit measures, the model demonstrates strong explanatory power. The adjusted R^2 is 0.62, indicating that our model explains approximately 62% of the variance in final exam scores. Furthermore, the RMSE is 10.30, suggesting a relatively high precision in predicting student outcomes compared to the standard deviation of the final exam (16.63).

Column 4 explores potential nonlinear effects by categorizing students based on their quiz performance relative to the mean. Using the lowest-performing group ($\leq$ Mean $-$ 1SD) as the reference, we find that the positive impact of quizzes is strictly monotonic. Students in the "middle-low" group (Mean $-$ 1SD $< PQ \leq$ Mean) score 6.10 points higher ($p < 0.01$), those in the "hiddle-high" group (Mean $< PQ \leq$ Mean $+$ 1SD) score 10.10 points higher ($p < 0.05$), and those in the highest performance bracket ($>$ Mean $+$ 1SD) exhibit a massive 16.25-point advantage ($p < 0.01$). This trend underscores that consistent engagement, as proxied by quiz performance, translates directly and increasingly into superior learning outcomes.

4.3 Sensitivity Analysis

To validate the reliability of our main findings, we performed a series of robustness checks using alternative specifications. Table 4 summarizes these results. First, we examined whether the grading policy itself drove the results. Column 1 uses the average of the top 2 quiz scores—matching the actual grading criteria—while column 2 uses the average of the remaining bottom 8 quizzes (those effectively dropped from the grade). Remarkably, the coefficient for the bottom 8 quizzes is even larger ($\beta = 7.09$, $p < 0.01$) than that of the top 2 ($\beta = 5.65$, $p < 0.01$). This is a critical finding: It indicates that students who maintained effort even in quizzes that did not count toward their final grade achieved higher exam scores. This strongly supports the preparation effect hypothesis, suggesting that the habit of continuous study, rather than the immediate grade incentive of a specific quiz, drives learning outcomes.

Table 4. Sensitivity analysis

	1	2	3	4	5
	Y	Y	Y	ln(Y)	Letter grade
Pop Quiz	5.65***	7.09***	8.96***	0.13***	1.70***
	(1.59)	(1.71)	(0.87)	(0.03)	(0.57)
Assignments	0.35	0.99	1.57	0.02	0.90***
	(0.88)	(0.63)	(0.99)	(0.02)	(0.20)
Attendance	0.26	−0.79**	0.14	0.00	0.54***
	(0.29)	(0.31)	(0.32)	(0.00)	(0.06)
Midterm	0.59***	0.49***	0.47***	0.01***	0.12***
	(0.06)	(0.10)	(0.03)	(0.00)	(0.02)
Academic controls	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	Yes	Yes	Yes	Yes
Section-by-semester FE	Yes	Yes	Yes	Yes	Yes
Cluster-robust SE	Yes	Yes	Yes	Yes	Yes
No. of obs.	257	257	257	257	257
R^2	0.60	0.63	0.65	0.62	0.53
Adj. R^2	0.56	0.59	0.62	0.59	
Root mean square error (RMSE)	10.98	10.60	10.16	0.16	

Notes: Columns 1 and 2 test alternative incentive structures: Column 1 uses the average of the "top 2" (graded) quizzes and assignments, while column 2 uses the average of the "bottom 8" quizzes. Column 3 clusters standard errors by students' residential origin. Column 4 uses the natural logarithm of the final exam score as the dependent variable. Column 5 employs an ordered logistic regression model with the final letter grade as the dependent variable. All specifications include full demographic and academic controls, along with section-by-semester fixed effects. McFadden's pseudo R^2 is reported for column 5. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Subsequent columns confirm the robustness of our estimates across different modeling choices. Column 3 clusters standard errors by students' residential origin (city/province) instead of major. This specification addresses potential spatial autocorrelation or regional heterogeneity in student backgrounds (Moulton 1990; Cameron and Miller 2015). The positive impact of quizzes remains robust and significant ($\beta = 8.96$, $p < 0.01$).

Column 4 employs a log-transformed dependent variable ($ln(FinalExam)$). Following the standard interpretation for semi-logarithmic equations (Halvorsen and Palmquist 1980), the coefficient ($\beta = 0.13$, $p < 0.01$) implies that a 1-point increase in quiz scores is associated with approximately a 13% increase in final exam scores, reinforcing the magnitude of the effect.

Finally, column 5 employs an ordered logistic regression model, using the final letter grade (e.g., A, B, C. …, F) as the dependent variable. Since letter grades represent ordinal rankings rather than continuous values, this approach is econometrically more appropriate than OLS for categorical outcomes (Greene 2012). It is important to note that because this model is estimated via maximum likelihood estimation (MLE) rather than least squares, standard goodness-of-fit measures such as adjusted R^2 and RMSE are not applicable. Instead, we report McFadden's pseudo R^2 (= 0.53), which serves as a comparative measure of model fit (McFadden 1974). The results confirm that higher quiz performance significantly increases the probability of achieving a higher letter grade ($p < 0.01$), providing strong corroborating evidence for our main results.

For the OLS specifications (columns 1–4), the model fit remains strong and stable, with adjusted R^2 values ranging from 0.56 to 0.62 and RMSE values remaining consistently around 10.16–10.98. These metrics reinforce that our baseline results are not artifacts of a specific model specification.

4.4 Placebo Test

To differentiate the preparation effect of unannounced quizzes from general student ability or instructor-specific effects, we conducted a placebo test using data from a comparable course, Principles of Macroeconomics. This course, taught by a different instructor within the same department, utilized scheduled (announced) quizzes, as explicitly stated in its syllabus. If the positive effects observed in our main analysis are driven solely by the unannounced nature of the assessment (and the resulting constant preparation), we would expect the effects of announced quizzes to be less consistent or distinct.

Table 5 presents the regression results for this placebo analysis. In the baseline model (column 1), announced quizzes appear to have a positive correlation with final exam performance. However, this relationship proves to be highly unstable. When demographic controls are introduced (column 2), the statistical significance of announced quizzes vanishes ($\beta = 2.06$, $p > 0.10$). Although significance returns when academic controls are included and standard errors are clustered by major (column 3) or residential origin (column 4), the magnitude of the coefficient fluctuates considerably compared to the robust stability observed for unannounced quizzes in our main analysis (Table 3).

Crucially, these results must be interpreted with significant caution due to the limited sample size ($N = 14$). We observe unusually high goodness-of-fit measures, with adjusted R^2 values ranging from 0.71 to 0.85. While high R^2 values typically indicate strong explanatory power, in the context of such a small sample relative to the number of predictors, this is a classic indicator of model overfitting rather than true predictive validity (Babyak 2004; Harrell 2015). The extremely low degrees of freedom inflate the R^2 and standard errors, making the estimates unreliable and prone to Type M (magnitude) and Type S (sign) errors (Button et al. 2013).

Therefore, while the placebo test offers suggestive evidence that announced quizzes do not consistently predict exam performance across all specifications, we refrain from drawing strong causal inferences from this auxiliary analysis. The contrast, however, highlights the relative robustness of our main findings: unlike the volatile estimates for announced quizzes, the impact of unannounced quizzes remains positive, significant, and stable across a large sample ($N = 257$) and rigorous specifications.

Table 5. Placebo test: The effect of announced quizzes

	1 No Controls	2 Demographic Controls	3 Academic Controls	4 Academic Controls
Announced Quiz	8.28***	2.06	10.82***	10.82
	(1.88)	(3.25)	(0.71)	(6.03)
Assignments	−4.36	−8.52	−6.69	−6.69
	(4.37)	(4.83)	(4.68)	(4.78)
Attendance	3.63	6.59**	3.43	3.43
	(3.97)	(2.54)	(1.96)	(2.73)
Midterm	0.46*	0.54***	0.35***	0.35*
	(0.22)	(0.13)	(0.08)	(0.13)
Academic controls	No	No	Yes	Yes
Demographic controls	No	Yes	No	No
Section-by-semester FE	No	No	No	No
Cluster-robust SE	Yes	Yes	Yes	Yes
No. of obs.	14	14	14	14
R^2	0.80	0.93	0.95	0.95
Adj. R^2	0.71	0.82	0.85	0.85
Root mean square error (RMSE)	6.23	4.83	4.44	4.44

Notes: This table reports the results of the placebo test using announced quizzes. The dependent variable is the final exam score. Column 1 presents baseline estimates without controls. Column 2 includes demographic controls. Columns 3 and 4 include academic controls. Standard errors in column 3 are clustered by academic major, while standard errors in column 4 are clustered by students' residential origin. Robust standard errors are reported in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5 Discussion

This section interprets the empirical findings presented above, focusing on the behavioral mechanisms that drive student performance. We first examine how unannounced quizzes shift study habits from episodic cramming to distributed practice, thereby supporting the preparation effect. Subsequently, we analyze the heterogeneous responses to the best-of-N incentive structure, highlighting the trade-off between anxiety reduction and strategic slacking among different student groups.

5.1 Mechanisms of the Preparation Effect: From Cramming to Distributed Practice

To further investigate these behavioral mechanisms, we categorized students into three groups based on their quiz performance: those who never achieved a perfect score ("none"), those who achieved a perfect score once ("once"), and those who scored perfect grades more than twice ("more than twice"). Figure A1 in the appendix presents the progression of quiz grades across the ten quizzes for each group.

The mechanism driving this effect can be understood through the lens of behavioral economics. Students often exhibit time-inconsistent preferences, postponing effort until immediately before a deadline—a phenomenon known as cramming or massed practice (Ariely and Wertenbroch 2002). While cramming may yield short-term performance spikes, it is less effective for long-term retention. Unannounced quizzes disrupt this procrastination cycle by introducing a "threat" of immediate evaluation, thereby nudging students toward distributed practice, which is widely cited as superior for memory retention (Cepeda et al. 2006; Roediger and Karpicke 2006).

Ideally, the best-of-*N* grading policy (counting only the top 2 scores) struck a delicate balance. It provided enough incentive to induce preparation while functioning as a safety net that reduced the paralyzing effects of high-stakes anxiety. By keeping quizzes low stakes (capped at 20% weight), the aim was to encourage students to treat them as ongoing opportunities for retrieval practice rather than as high-pressure evaluations (Black and Wiliam 1998). Our empirical findings suggest that the motivational mechanism of unannounced quizzes derives less from their absolute grade weight and more from the uncertainty of timing. In this sense, the preparation effect is a behavioral response to unpredictability, inducing students to maintain a "readiness potential" throughout the semester (Kamuche 2007).

5.2 Strategic Incentives and Heterogeneous Responses

To further investigate these behavioral mechanisms, we analyzed student performance patterns based on their accumulation of perfect scores. This analysis distinguishes between students who satisfied the grading criteria early versus those who persisted throughout the semester.

Figure A2 in the appendix categorizes students based on the number of perfect scores achieved before the midterm (quizzes 1–5). The "high achievers (Pre-midterm)" group—students who had already secured two or more perfect scores by the midpoint—exhibits a significant declining trend in quiz grades during the second half of the semester. This behavior is consistent with the economic theory of satisficing, first proposed by Simon (1956). Since these students had already maximized their grade potential under the best-of-*N* policy, the marginal benefit of preparing for subsequent quizzes dropped to zero. Consequently, they engaged in rational optimization of their time resources, reallocating effort away from quizzes to other pursuits (Becker 1965). While this confirms that students respond efficiently to incentives, it highlights a trade-off: The policy successfully reduced anxiety but unintentionally dampened the motivation for continuous preparation among top-performing students once they felt "safe" (Bloom 1968).

In contrast, the "none" and "once" groups in Figure A1 (categorized by total semester performance) display noticeable upticks in quiz grades after quiz 3 and, particularly, after the midterm. This trend suggests that students who performed poorly on the midterm exam subsequently increased their study efforts to compensate. For these students, the best-of-*N* policy acted as a continuous incentive mechanism. Unlike cumulative grading systems where early failures permanently damage a student's grade-point average (GPA), the "drop-lowest" feature ensured that the marginal benefit of an additional high score remained positive until the very end. This structure aligns with the principles of mastery learning, allowing students to recover from early setbacks without being penalized, thereby sustaining motivation for those who need it most (Guskey 2007).

Overall, these dynamics highlight a critical nuance in assessment design. The policy balances fairness and flexibility but creates distinct behavioral incentives: strategic slacking for early achievers and sustained persistence for those seeking recovery. These findings resonate with the broader formative assessment literature, which emphasizes that while flexible grading can lower stress, it must be carefully designed to sustain long-term engagement across all learner profiles (Black and Wiliam 1998; Wiliam 2011).

6 Conclusion

This study provides empirical evidence supporting the efficacy of unannounced quizzes as a pedagogical tool in introductory economics courses. Utilizing a dataset of 257 students and controlling for section-by-semester fixed effects, we find that performance in unannounced quizzes is a robust predictor of final exam achievement. Specifically, a 1-point increase in the average quiz score is associated with an 8.67-point increase in the final exam score.

Our findings offer three key insights into the mechanics of student assessment. First, the positive impact of quizzes appears to be driven by the preparation effect rather than simple grade incentives. The robustness check revealed that performance in quizzes excluded from the final grade (the “bottom 8”) was a stronger predictor of exam success than the graded quizzes (the “top 2”). This suggests that the uncertainty of unannounced assessments successfully nudged students toward distributed practice and continuous engagement, mitigating the tendency for procrastination and cramming often associated with scheduled exams.

Second, the best-of-N grading policy (counting only the top 2 scores) presents a critical trade-off between anxiety reduction and motivation. While this policy successfully functioned as a safety net—encouraging lower-performing students to persist and recover from early setbacks—it inadvertently induced strategic slacking or satisficing behavior among high achievers. As predicted by economic models of rational time allocation (Simon 1956; Becker 1965), high-performing students reduced their effort once they secured the maximum possible credit. This highlights the need for nuanced course designs that protect struggling students without capping the incentives for top performers.

Third, while our results are robust across various specifications, we acknowledge important limitations. The placebo test utilizing announced quizzes relied on a small sample ($N = 14$), which, despite suggestive contrasts, limits our ability to draw definitive causal comparisons between announced and unannounced formats. Future research should aim to replicate this study across multiple institutions and instructors to enhance external validity and further disentangle the effects of assessment timing from instructor-specific characteristics. Finally, the exclusion of 13 students who missed the final exam may introduce survival bias, as those unable to cope with the quiz burden might have dropped out, potentially leading to an overestimation of the preparation effect.

In conclusion, unannounced quizzes are a potent instrument in the economics instructor’s toolkit. When implemented with a flexible grading policy, they can effectively foster a habit of continuous preparation. However, instructors must carefully calibrate the incentive structure to minimize anxiety while sustaining motivation across the entire spectrum of student abilities.

About the Authors: Moon Joon Kim (mkim238@gmu.edu), Department of Economics, George Mason University Korea.

Acknowledgments: This research was supported by funding from the 2026 Mason Korea Faculty Research and Development Award program.

Human Subject Information: This study was conducted in accordance with ethical standards for research involving human subjects. Because the analysis relies on de-identified retrospective educational records, the Institutional Review Board (IRB) of George Mason University determined that the study was exempt from formal review (Protocol #[STUDY00000074]).

Appendix

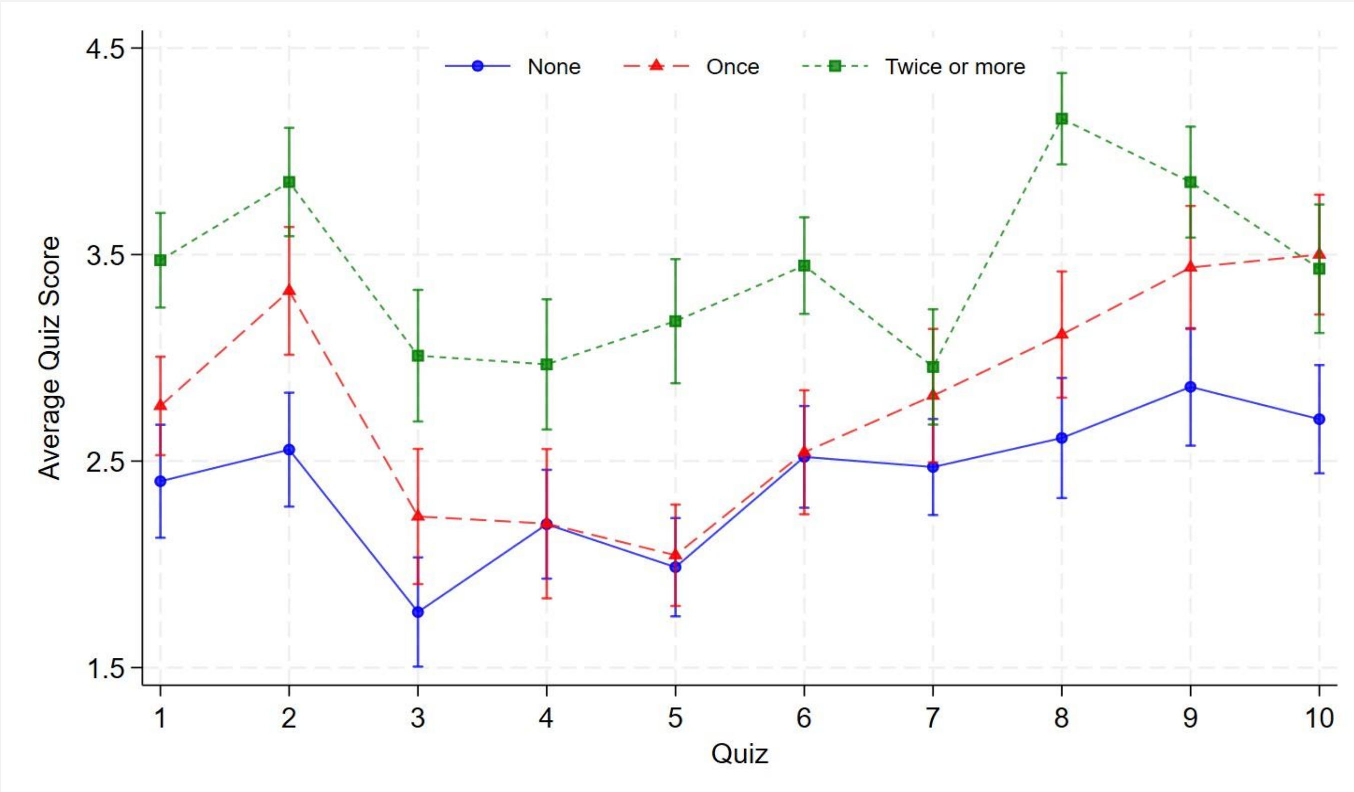

Figure A1. Trends in quiz performance by total semester achievement

Note: The horizontal axis represents the chronological order of the ten unannounced quizzes. The vertical axis denotes the average quiz score on a 0–5 scale. Students are categorized into three groups based on the *total* number of perfect scores achieved throughout the semester: "none" (0 perfect scores), "once" (1 perfect score), and "twice or more" (2 or more perfect scores).

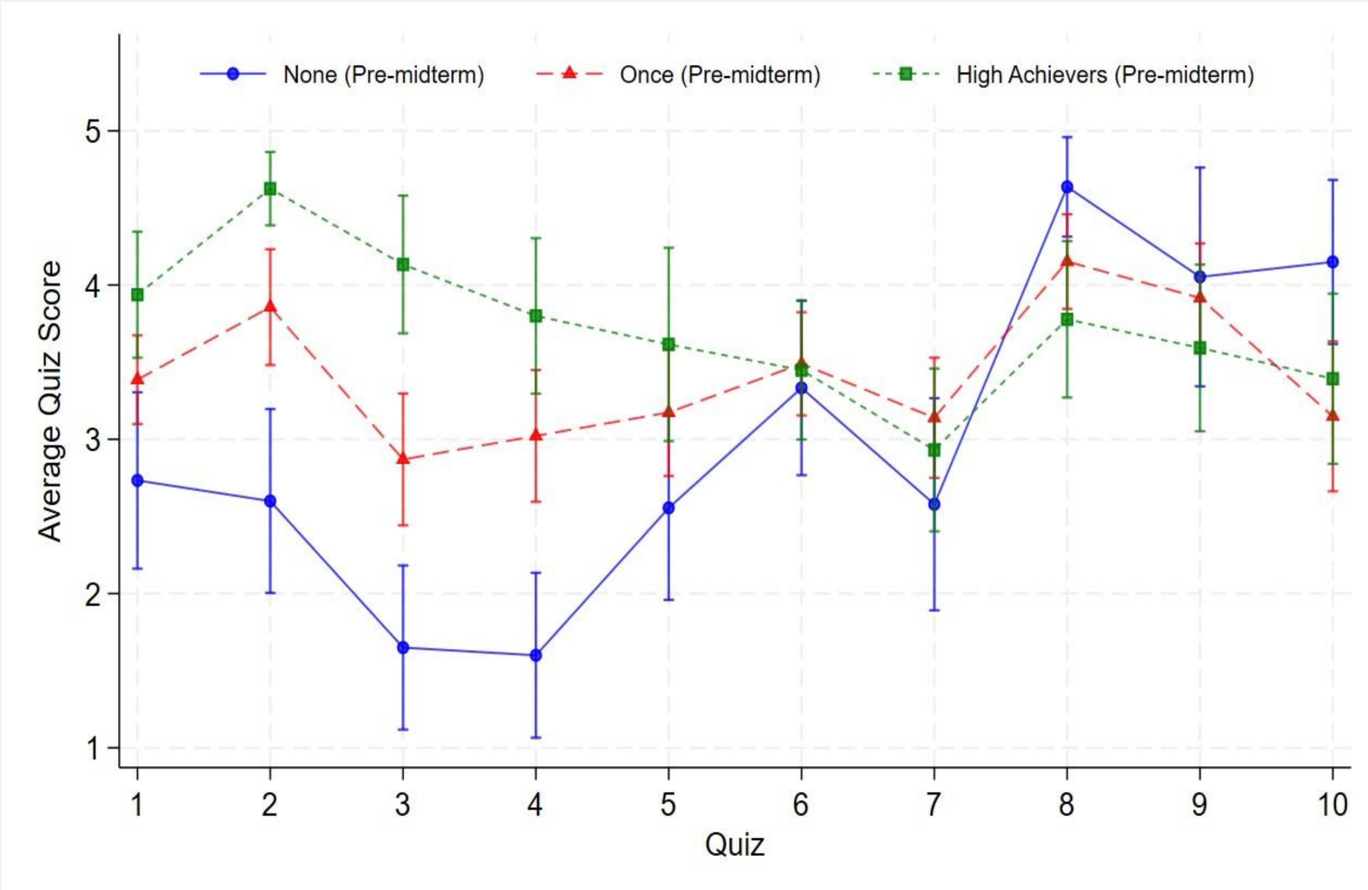

Figure A2. Trends in quiz performance by early-semester achievement (pre-midterm)

Note: The horizontal axis represents the chronological order of the ten unannounced quizzes. The vertical axis denotes the average quiz score on a 0–5 scale. Students are categorized into three groups based on the number of perfect scores achieved *before the midterm exam* (quizzes 1–5): "none (pre-midterm)" (0 perfect scores), "once (pre-midterm)" (1 perfect score), and "high achievers (pre-midterm)" (2 or more perfect scores).

References

Andrews, D.W.K. 2005. "Cross-Section Regression with Common Shocks." *Econometrica* 73(5):1551–1585. https://doi.org/10.1111/j.1468-0262.2005.00629.x

Ariely, D., and K. Wertenbroch. 2002. "Procrastination, Deadlines, and Performance: Self-Control by Precommitment." *Psychological Science* 13(3):219–224. https://doi.org/10.1111/1467-9280.00441

Azorlosa, J. L. 2011. "The Effect of Announced Quizzes on Exam Performance: II." *Journal of Instructional Psychology* 38(1):3–7.

Babyak, M.A. 2004. "What You See May Not Be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models." *Psychosomatic Medicine* 66(3):411–421. https://doi.org/10.1097/01.psy.0000127692.23278.a9

Bangert-Drowns, R.L., C.-L.C. Kulik, and J.A. Kulik. 1991. "Effects of Frequent Classroom Testing." *Journal of Educational Research* 85(2):89–99. https://doi.org/10.1080/00220671.1991.10702818

Becker, G.S. 1965. "A Theory of the Allocation of Time." *Economic Journal* 75(299):493–517. https://doi.org/10.2307/2228949

Bjork, E.L., and R.A. Bjork. 2011. "Making Things Hard on Yourself, but in a Good Way: Creating Desirable Difficulties to Enhance Learning." In M.A. Gernsbacher, R.W. Pew, L.M. Hough, and J.R. Pomerantz, eds. *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*. Worth Publishers, 56–64.

Bjork, R.A. 1994. "Memory and Metamemory Considerations in the Training of Human Beings." In J. Metcalfe and A.P. Shimamura, *Metacognition: Knowing about Knowing*. MIT Press, 185–205.

Bjork, R.A., J. Dunlosky, and N. Kornell. 2013. "Self-Regulated Learning: Beliefs, Techniques, and Illusions." *Annual Review of Psychology* 64:417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Black, P., and D. Wiliam. 1998. "Assessment and Classroom Learning." *Assessment in Education: Principles, Policy & Practice* 5(1):7–74. https://doi.org/10.1080/0969595980050102

Blasco-Arcas, L., I. Buil, B. Hernández-Ortega, and F.J. Sese. 2013. "Using Clickers in Class. The Role of Interactivity, Active Collaborative Learning and Engagement in Learning Performance." *Computers & Education* 62:102–110. https://doi.org/10.1016/j.compedu.2012.10.019

Bloom, B.S. 1968. "Learning for Mastery." *Evaluation Comment* 1(2):1–12.

Boud, D. 2000. "Sustainable Assessment: Rethinking Assessment for the Learning Society." *Studies in Continuing Education* 22(2):151–167. https://doi.org/10.1080/713695728

Button, K.S., J.P.A. Ioannidis, C. Mokrysz, B.A. Nosek, J. Flint, E.S.J. Robinson, and M.R. Munafò. 2013. "Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience." *Nature Reviews Neuroscience* 14(5):365–376. https://doi.org/10.1038/nrn3475

Cameron, A.C., and D.L. Miller. 2015. "A Practitioner's Guide to Cluster-Robust Inference." *Journal of Human Resources* 50(2):317–372. https://doi.org/10.3368/jhr.50.2.317

Carpenter, S.K., N.J. Cepeda, H. Pashler, E. Vul, and J.T. Wixted. 2008. "The Effects of Tests on Learning and Forgetting." *Memory & Cognition* 36(2):438–448. https://doi.org/10.3758/MC.36.2.438

Cassady, J.C., and R.E. Johnson. 2002. "Cognitive Test Anxiety and Academic Performance." *Contemporary Educational Psychology* 27(2):270–295. https://doi.org/10.1006/ceps.2001.1094

Cepeda, N.J., H. Pashler, E. Vul, J.T. Wixted, and D. Rohrer. 2006. "Distributed Practice in Verbal Recall Tasks: A Review and Quantitative Synthesis." *Psychological Bulletin* 132(3):354–380. https://doi.org/10.1037/0033-2909.132.3.354

Chatterjee, S., and A.S. Hadi. 2012. *Regression Analysis by Example*, 5th ed. Wiley.

Eccles, J.S., and A. Wigfield. 2002. "Motivational Beliefs, Values, and Goals." *Annual Review of Psychology* 53:109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Graham, R.B. 1999. "Unannounced Quizzes Promotes Academic Performance." *Psychological Reports* 84(1):267–273. https://doi.org/10.2466/pr0.1999.84.1.267

Greene, W.H. 2012. *Econometric Analysis*, 7th ed. Pearson Education.

Guskey, T.R. 2007. "Closing Achievement Gaps: Revisiting Benjamin S. Bloom's 'Learning for Mastery'." *Journal of Advanced Academics* 19(1):8–31. https://doi.org/10.4219/jaa-2007-704

Halvorsen, R., and R. Palmquist. 1980. "The Interpretation of Dummy Variables in Semilogarithmic Equations." *American Economic Review* 70(3):474–475.

Harrell, F.E. 2015. Re*gression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*, 2nd ed. Springer. https://doi.org/10.1007/978-3-319-19425-7

Imbens, G.W., and J.M. Wooldridge. 2009. "Recent Developments in the Econometrics of Program Evaluation." *Journal of Economic Literature* 47(1):5–86. https://doi.org/10.1257/jel.47.1.5

Kamuche, F.U. 2007. "The Effects of Unannounced Quizzes on Student Performance: Further Evidence." *College Teaching Methods & Styles Journal* 3(2):21–26. https://doi.org/10.19030/ctms.v3i2.5277

Karpicke, J.D., A.C. Butler, and H.L. Roediger. 2009. "Metacognitive Strategies in Student Learning: Do Students Practice Retrieval When They Study on Their Own?" *Memory* 17(4):471–479. https://doi.org/10.1080/09658210802647009

Karpicke, J.D., and H.L. Roediger. 2008. "The Critical Importance of Retrieval for Learning." *Science* 319(5865):966–968. https://doi.org/10.1126/science.1152408

Latif, E., and S. Miles. 2020. "The Impact of Assignments and Quizzes on Exam Grades: A Difference-in-Difference Approach." *Journal of Statistics Education* 28(3):289–294. https://doi.org/10.1080/10691898.2020.1807366

Leeming, F.C. 2002. "The Exam-A-Day Procedure Improves Performance in Psychology Classes." *Teaching of Psychology* 29(3):210–212. https://doi.org/10.1207/S15328023TOP2903_06

Mawhinney, V.T., D.E. Bostow, D.R. Laws, G.J. Blumenfeld, and B.L. Hopkins. 1971. "A Comparison of Students Studying-Behavior Produced by Daily, Weekly, and Three-Week Testing Schedules." *Journal of Applied Behavior Analysis* 4(4):257–264. https://doi.org/10.1901/jaba.1971.4-257

McFadden, D. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In P. Zarembka, ed. *Frontiers in Econometrics*. Academic Press, pp. 105–142.

Moulton, B.R. 1990. "An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables on Micro Units." *Review of Economics and Statistics* 72(2):334–338. https://doi.org/10.2307/2109724

Putwain, D.W. 2008. "Deconstructing Test Anxiety." *Emotional and Behavioural Difficulties* 13(2):141–155. https://doi.org/10.1080/13632750802027631

Putwain, D.W., and W. Symes. 2011. "Teachers' Use of Fear Appeals in the Mathematics Classroom: Worrying or Motivating Students?" *British Journal of Educational Psychology* 81(3):456–474. https://doi.org/10.1348/2044-8279.002005

Roediger, H.L., and J.D. Karpicke. 2006. "Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention." *Psychological Science* 17(3):249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Romer, D. 1993. "Do Students Go to Class? Should They?" *Journal of Economic Perspectives* 7(3):167–174. https://doi.org/10.1257/jep.7.3.167

Segool, N.K., J.S. Carlson, A.N. Goforth, and N. von der Embse. 2013. "Heightened Test Anxiety among Young Children: Elementary School Students' Anxious Responses to High-Stakes Testing." *Psychology in the Schools* 50(5):489–499. https://doi.org/10.1002/pits.21689

Simon, H.A. 1956. "Rational Choice and the Structure of the Environment." *Psychological Review* 63(2):129–138. https://doi.org/10.1037/h0042769

Stanca, L. 2006. "The Effects of Attendance on Academic Performance: Panel Data Evidence for Introductory Microeconomics." *Journal of Economic Education* 37(3):251–266. https://doi.org/10.3200/JECE.37.3.251-266

Stowell, J.R. 2015. "Online Open-Book Testing in Face-to-Face Classes." *Scholarship of Teaching and Learning in Psychology* 1(1):7–13. https://doi.org/10.1037/stl0000013

Walstad, W.B. 2001. "Improving Assessment in University Economics." *Journal of Economic Education* 32(3):281–294. https://doi.org/10.1080/00220485.2001.10844715

Walstad, W.B., and W.E. Becker. 1994. "Achievement and Attitude Measurement in Economics." In P. Saunders and W.B. Walstad, eds. *The Principles of Economics Course: A Handbook for Instructors*. McGraw-Hill, pp. 163–183.

Wiliam, D. 2011. "What Is Assessment for Learning?" *Studies in Educational Evaluation* 37(1):3–14. https://doi.org/10.1016/j.stueduc.2011.03.001

Wooldridge, J.M. 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. MIT Press.

———. 2015. *Introductory Econometrics: A Modern Approach*, 6th ed. South-Western Cengage Learning.

Yerkes, R.M., and J.D. Dodson. 1908. "The Relation of Strength of Stimulus to Rapidity of Habit-Formation." *Journal of Comparative Neurology and Psychology* 18(5):459–482. https://doi.org/10.1002/cne.920180503

Zeidner, M. 1998. *Test Anxiety: The State of the Art*. Plenum Press.

Zimmerman, B.J. 2002. "Becoming a Self-Regulated Learner: An Overview." *Theory into Practice* 41(2):64–70. https://doi.org/10.1207/s15430421tip4102_2

DOI: https://doi.org/10.71162/aetr.628384