Beyond Repetition: The Role of Varied Questioning and Feedback in Knowledge Generalization

Gautam Yadav
Carnegie Mellon University
Pittsburgh, PA, USA
gyadav@andrew.cmu.edu

Paulo F. Carvalho
Carnegie Mellon University
Pittsburgh, PA, USA
pcarvalh@cs.cmu.edu

Elizabeth A. McLaughlin
Carnegie Mellon University
Pittsburgh, PA, USA
mimim@cs.cmu.edu

Kenneth R. Koedinger
Carnegie Mellon University
Pittsburgh, PA, USA
koedinger@cmu.edu

ABSTRACT

This study examines the effects of question type and feedback on learning outcomes in a hybrid graduate-level course. By analyzing data from 32 students over 30,198 interactions, we assess the efficacy of unique versus repeated questions and the impact of feedback on student learning. The findings reveal students demonstrate significantly better knowledge generalization when encountering unique questions compared to repeated ones, even though they perform better with repeated opportunities. Moreover, we find that the timing of explanatory feedback is a more robust predictor of learning outcomes than the practice opportunities themselves. These insights suggest that educational practices and technological platforms should prioritize a variety of questions to enhance the learning process. The study also highlights the critical role of feedback; opportunities preceding feedback are less effective in enhancing learning.

CCS CONCEPTS

• **Applied computing** → **E-learning**; **Interactive learning environments**; • **Human-centered computing** → *Empirical studies in HCI*.

KEYWORDS

Knowledge Generalization, Learning Outcomes, Instructional Design, Repetition, Variability, Feedback, Student Modeling, Model Comparison

ACM Reference Format:
Gautam Yadav, Paulo F. Carvalho, Elizabeth A. McLaughlin, and Kenneth R. Koedinger. 2024. Beyond Repetition: The Role of Varied Questioning and Feedback in Knowledge Generalization. In *Proceedings of the Eleventh ACM Conference on Learning @ Scale (L@S '24), July 18–20, 2024, Atlanta, GA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3657604.3664688

L@S '24, July 18–20, 2024, Atlanta, GA, USA

ACM ISBN 979-8-4007-0633-2/24/07
https://doi.org/10.1145/3657604.3664688

1 INTRODUCTION

In online education, immediate feedback with practice questions significantly enhances learning outcomes [11]. A recent study [9] found that students typically require about seven practice opportunities with feedback to master a knowledge component to an 80% correctness level, emphasizing the importance of practice beyond initial instruction. Despite the benefits, creating effective assessments with feedback is both time-consuming and costly [3, 5]. One approach to providing more practice involves asking students to repeat assessments once the question pool is exhausted, continuing until they achieve the desired performance for each knowledge component. To illustrate the type of unique questions used in the course, Figure 1 presents three unique questions mapped to the same knowledge component, highlighting the variability in question content aimed at enhancing learning generalization.

Firstly, we examine the knowledge component (KC) models to determine whether the timing of explanatory feedback is a more reliable predictor of student learning than practice opportunity. A mapping of KCs to problem steps constitutes a KC Model. Such a model typically utilizes log data to track changes in student performance as the count of opportunities (i.e., interactions or practices with a given KC) increases [16]. In our implementation, feedback timing differs between inline and quiz assessments: inline provides immediate feedback following the assessment, whereas quizzes provide feedback for all assessment items collectively after the student has submitted the quiz.

After establishing the best-fit model, central to our investigation is the comparative efficacy of unique and repeated questions. Variability in practice opportunities is presumed to hinder immediate performance but enhance learning and transfer [15]. This research, conducted within a hybrid graduate-level course, seeks to determine whether the repetition of assessments enhances responses to subsequent, unseen assessments for the same knowledge component. Understanding the best strategies for their development can significantly impact the scalability and effectiveness of online courses.

We also address the frequency of question repetition and its correlation with overall student performance, seeking a threshold where the benefits of repetition decrease, potentially detracting from learning effectiveness.

Our research questions are:

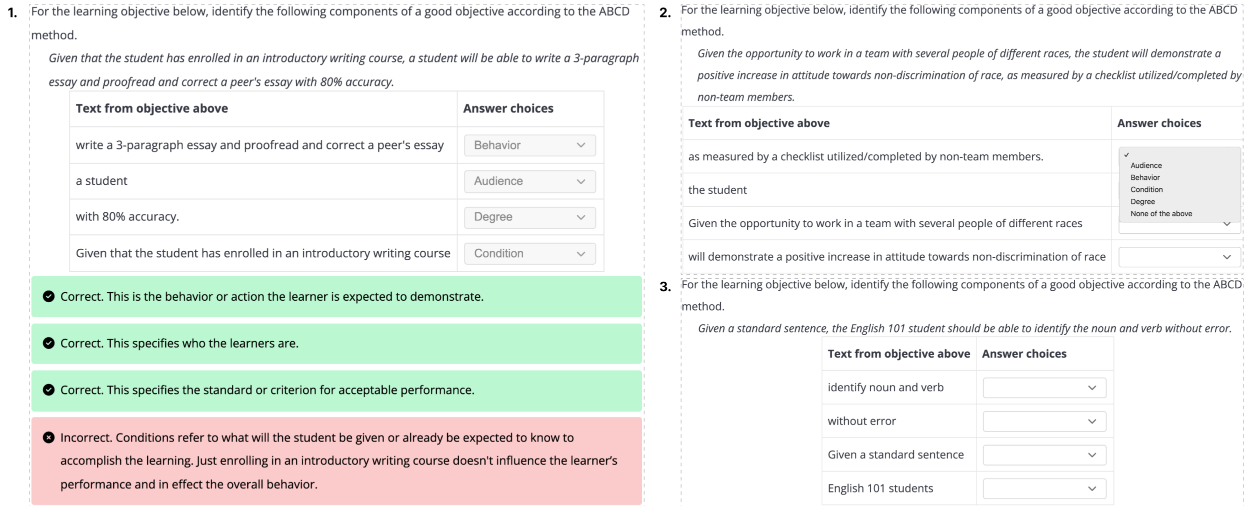

Figure 1: Three unique questions associated with the same knowledge component. Feedback for the first is shown as an example.

(1) RQ1: Does learning occur more effectively when feedback is provided after the completion of a quiz, rather than during the quiz as each practice opportunity is attempted?
(2) RQ2: How do unique and repeated opportunities influence students' ability to generalize knowledge to unseen items?
(3) RQ3: How does the frequency of repeated question attempts correlate with students' overall performance?

2 RELATED WORK

Previous studies demonstrate the prevalence of a 'doer effect,' where active practice significantly enhances student learning outcomes, surpassing passive learning methods, by comparing model fits from data available in DataShop [8–10].

The effect of repetition including identical instruction on learning outcomes has classically been studied mainly based on memorization (e.g., word lists) [12, 13]. The role of repetition in learning, especially the contrast between identical task repetition and varying task repetition, has been explored in various contexts but remains underexplored in terms of knowledge generalization. For instance, research in motor skill acquisition in virtual reality settings for children with Developmental Coordination Disorder (DCD) indicated that the type of practice either repetitive versus variable did not significantly affect the transfer of learned skills to real-world tasks [2]. Similarly, studies on language acquisition suggest that while identical task repetition may improve performance in structured tests, it does not significantly outperform task-type repetition in terms of long-term knowledge retention [17].

Knowledge component modeling and learning curve analysis have been used to evaluate student learning in several domains of higher education, such as psychology, biology, statistics, and programming [1, 14]. Prior research exploring KC models has focused on course improvement [14] or discovering the best model through the Additive Factors Model [6].

The timing of feedback has also been a critical area of research. Studies examining immediate versus delayed feedback across multiple college classes found no significant differences in learning outcomes between the two at scale [7]. This finding is particularly relevant to our study as it suggests that if delayed feedback were consistently more effective, the opportunities before feedback would not be as critical for learning.

This study aims to contribute to this ongoing discussion by specifically examining how identical versus different task repetition impacts knowledge generalization. Additionally, we explore alternate KC model mapping based on the timing of feedback, an area that has not been extensively studied.

3 METHODS

3.1 Participants and Data Collection

This study analyzed data from 32 students in a hybrid graduate-level course, "E-Learning Design Principles and Methods" [8] offered through the Open Learning Initiative platform. The course consists of 20 modules with interactive activities, complemented by pre- and post-quizzes. The classroom/remote portion of the course includes lectures, four exams, and two projects. Prior to analysis, we excluded data from students who dropped out of the course. Additionally, opportunities not associated with a specific knowledge component (KC) or labeled with an "unknown" outcome were removed, ensuring a dataset of 30,198 opportunities across 78 KCs.

The interactions with course materials termed "opportunities" include a wide range of activities, from inline formative assessments to review practice quizzes. Student responses to these tasks are automatically tagged as correct when students answer correctly on their first attempt without asking for a hint. Otherwise, the task response is tagged as incorrect. Only the student's first attempt is considered to estimate performance at a given task opportunity, though subsequent student attempts and system feedback are important contributors to learning. We define learning as a positive change in performance and operationalize learning as a reduction in error rate (or increase in correctness rate) over successive opportunities to perform a task associated with a specific knowledge component [9].

Each module had both a pre-quiz before instruction and a post-quiz, after instruction. Each quiz typically comprised around 8 to 10 selected-response questions drawing from a common pool of questions. Inline activities were specifically designed to mirror the content of quiz and exam questions closely, aligning with specific learning objectives and providing for extra practice. Each quiz had a fixed number of questions for each knowledge component based on their importance as determined by the instructor. Students were permitted to retake post-quiz assessments multiple times, with the highest score across all quiz attempts counted towards their final grade. Students received explanatory feedback on both correct and incorrect responses immediately for inline activities and after quiz completion for each quiz question .

3.2 Labeling Process

To precisely measure learning opportunities, we distinguished between opportunities based on each attempt a student made and opportunities adjusted for quiz feedback. We determined the better KC model based on the Akaike Information Criterion (AIC) [16].

In the **Practice Opportunity Labeling** approach, every practice attempt is counted as a separate learning opportunity. This method allows us to track each interaction a student has with the material, classified as either a unique opportunity or a repeat opportunity, based on the sequence of attempts made for a particular KC. Table 1 provides an example. It illustrates the sequence of learning opportunities a student received for different knowledge components kc1 and kc2 while answering questions (q1-q5) on different quizzes (z1-z3). Focusing on the Practice columns, we see in the second row one prior learning opportunity (Practice Opp=1) and because it was a different question it was a unique opportunity (Unique=1). Skipping to the fourth row (q3), we see 3 prior opportunities (Opp=3) with two that were unique (q1 and q2 in rows 1 and 2) and one that was a repeat (q1 a second time in row 3).

The **Quiz-Adjusted Opportunity Labeling** columns count opportunities only after a quiz is complete and feedback has been provided on all questions. Thus, we see in row 2, that there are no prior learning opportunities yet (Quiz-Adjusted Opp=0) because the student has yet to receive feedback on their answer to question q1. This adjustment provides insights into how the timing of feedback impacts the learning progression, by only considering those interactions that are reinforced by feedback. For inline questions, the labeling is the same as the Practice Opportunity because immediate feedback is given after each individual assessment.

3.3 Data Analysis

For RQ1, we conducted a learning curve analysis to categorize knowledge components (KCs) based on their learning outcomes over successive task opportunities (using DataShop's default thresholds [4]). This method helped us identify distinct learning patterns across KCs by examining changes in error rates. For RQ2, we utilized a generalized linear mixed model (GLMM) to assess the impact of question repetition and feedback timing on student learning outcomes. The dependent variable in this model is 'First.Attempt,' indicating whether a student's first attempt at a question was correct (coded as 1) or incorrect (coded as 0). This model incorporated

Table 1: Practice Opportunity and Quiz-Adjusted Opportunity Labeling

			Practice			Quiz-Adjusted		
Question	KC	Quiz	Opp	Unique	Repeat	Opp	Unique	Repeat
q1	kc1	z1	0	0	0	0	0	0
q2	kc1	z1	1	1	0	0	0	0
q1	kc1	z2	2	1	1	2	1	1
q3	kc1	z2	3	2	1	2	1	1
q2	kc1	z3	4	2	2	4	2	2
q4	kc2	z1	0	0	0	0	0	0
q5	kc2	z1	1	1	0	0	0	0
q4	kc2	z2	2	1	1	2	1	1

fixed effects for each type of question repetition—unique, same-day repeated, and different-day repeated—and random effects to account for individual differences in student prior knowledge (intercept) and in KC difficulty (intercept) and learning rate (slope). For RQ3, we analyzed the frequency of question repetitions and its correlation with overall student performance using correlation and regression analyses.

4 RESULTS

We classified KCs into distinct categories based on learning curve analysis setting a student threshold at 10 to filter out less informative data points in DataShop:

- 'Low and Flat' [1 KC] includes curves where all points are below a 20% error rate, indicating consistently high performance from the onset.
- 'No Learning' [2 KCs] identifies curves with a slope below the 0.001 AFM slope threshold, signifying no significant improvement in student performance over time.
- 'Still High' [6 KCs] encompasses curves where the final error rate remains above the 40% high error threshold, suggesting that students have not reached a satisfactory level of understanding.
- 'Good' [69 KCs] denotes curves with a significant positive slope, illustrating that effective learning is occurring as students improve with more opportunities.

All KCs met the minimum opportunity threshold of three, so we did not assign any to the 'Too Little Data' category.

4.1 RQ1: Learning Occurs After Feedback

The quiz-adjusted model in Table 2, which only counts opportunities after quiz feedback is provided, fitted the data better (AIC = 35126.2) than the model based on practice opportunity counts (AIC = 35165.4). This improvement indicates that counting opportunities only after quizzes—and thereby incorporating the effect of feedback—results in a more accurate model of student learning.

4.2 RQ2: Much Better Generalization from Varied Questions than Repeated Questions

The generalized linear mixed model (GLMM) analysis, using the Quiz-Adjusted model from Table 2, was conducted after finding the

Table 2: Comparison of KC models based on different labeling approaches (practice-based vs quiz-adjusted)

Labeling	AIC	BIC	logLik	deviance	df.resid
Practice-Based	35165.4	35248.6	-17572.7	35145.4	30188
Quiz-Adjusted	35126.2	35209.3	-17553.1	35106.2	30188

Table 3: Model parameter estimates

Parameter	Coef.	Std. Error	z	Pr(>|z|)
(Intercept)	0.413	0.101	4.101	<.001
unique	0.092	0.013	7.020	<.001
repeat: same day	0.006	0.003	1.851	.064
repeat: different day	-0.018	0.018	-1.006	.314

best-fit model. Table 3 provides compelling evidence for the superiority of varied questions over repeated ones in fostering student generalization abilities on unseen items. Unique learning opportunities showed a significant positive effect on students' ability to generalize knowledge to unseen items ($p < .001$). This result indicates that students benefit considerably more from engaging with new material, supporting the hypothesis that varied questions significantly enhance learning generalization.

For same-day repeated opportunities, the analysis showed a marginally positive effect ($p = .064$), suggesting that immediate repetition may enhance short-term learning. In contrast, different-day repeated opportunities exhibited a non-significant negative trend ($p = .314$). This indicates that the benefits of repetition for long-term retention are less clear.

4.3 RQ3: Impact of Repetition

We initially analyzed the difference in performance between unique and repeated attempts. Our analysis showed that students had an average performance of 64.59% on unique attempts, whereas on repeated attempts, they scored higher, at 71.31% ($p < .001$).

Next, we explored how the frequency of question repetitions correlates with the average quiz scores across the course. We found a significant negative correlation ($r = -0.59$, $p < .001$) indicating that students who tend to score lower grades on the quizzes frequently repeat questions. This result is consistent with the course policy that students can retake quizzes when their score is lower than desired such that lower scores yield more repetition.

5 DISCUSSION AND LIMITATIONS

The analysis indicates that unique learning opportunities positively impact performance and, at best, same-day repetition may offer marginal benefits. A key implication is that we should add more questions to our pool of questions so as to maximize practice on unique questions and minimize repetitions of identical questions. Indeed, we have been using generative AI to help us in adding new questions for a given knowledge component.

Another finding is that incrementing the count of learning opportunities per knowledge component *after* a quiz rather than *during* the quiz yields a better prediction of students' future performance. Because students get feedback after the quiz rather than during, this result provides further evidence for the value of feedback in learning – namely, while there is little or no improvement from question to question without feedback during a quiz, there is improvement revealed after the quiz and receiving feedback.

Our study's insights come with limitations. Specifically, we combined data from 78 distinct knowledge components, each with varying question pool size and repetition frequency. This amalgamation likely skews results, blending data from frequently repeated questions with those revisited after longer periods. Such an approach may mask the subtleties present when examining individual knowledge components, where repetition's impact on performance could vary significantly. Future research should disaggregate these effects and closely examine the impacts of question pool size and the timing of repetitions within each knowledge component to more precisely determine how these factors influence learning outcomes.

Another limitation is the correlational nature of our findings, as the relationship between unique opportunities and performance may not imply causation. An alternative explanation is that students who engage more with unique opportunities are already stronger learners, while those who repeat questions more frequently might be weaker learners. This explanation seems unlikely for two reasons. First, given our past observation of low variability in student learning rates [9] in online practice with feedback, it is unlikely there are particularly stronger and weaker learners. Second, given our analysis relies on within-student comparisons and controls for student prior knowledge, the result is unlikely based on student differences. Nevertheless, further investigation is warranted.

6 CONCLUSION

We presented a novel analysis, incrementing opportunity count within versus after a quiz, that provides further evidence for the importance of feedback during practice in aiding student learning. More importantly, we found evidence that engaging students with unique learning opportunities is correlated with higher future performance more so than engaging in repetitions of the same questions. While identical question repetition helps performance on the repeated question, it does little to aid generalization to unseen questions. This result is interesting scientifically in that theories of learning that emphasize the role of memory rather than a general concept or skill induction may be interpreted as predicting learning benefits from repeating the same question, whereas we found only a small, non-significant trend for such. Students experiencing repeated questions appear to memorize the answer verbatim (yielding better performance on repeated questions) but are much less likely to engage in attempts to induce a general concept or skill than when experiencing different questions tapping the same general concept or skill. An important practical implication is the value of having a larger pool of practice questions for each knowledge goal so as to facilitate the benefits of varied practice with unique questions and reduce the chance of repeating identical questions, which may do little more than take student time.

ACKNOWLEDGMENTS

We thank Hui Cheng for providing custom reports of the KC Model metrics and her programming expertise.

REFERENCES

[1] Olle Bälter, Dawn Zimmaro, and Candace Thille. 2018. Estimating the minimum number of opportunities needed for all students to achieve predicted mastery. *Smart Learning Environments* 5, 1 (2018), 15.

[2] Emmanuel Bonney, Lemke Dorothee Jelsma, Gillian D Ferguson, and Bouwien CM Smits-Engelsman. 2017. Learning better by repetition or variation? Is transfer at odds with task specific training? *PLoS One* 12, 3 (2017), e0174214.

[3] Sandra L Clifton and Cheryl L Schriner. 2010. Assessing the quality of multiple-choice test items. *Nurse educator* 35, 1 (2010), 12–16.

[4] DATASHOP TEAM. 2016. Learning Curve. https://pslcdatashop.web.cmu.edu/help?page=learningCurve Accessed: April 15, 2024.

[5] David DiBattista and Laura Kurzawa. 2011. Examination of the quality of multiple-choice items on classroom tests. *Canadian Journal for the Scholarship of Teaching and Learning* 2, 2 (2011), 4.

[6] Tomáš Effenberger, Radek Pelánek, and Jaroslav Čechák. 2020. Exploration of the robustness and generalizability of the additive factors model. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*. 472–479.

[7] Emily R Fyfe, Joshua R de Leeuw, Paulo F Carvalho, Robert L Goldstone, Janelle Sherman, David Admiraal, Laura K Alford, Alison Bonner, Chad E Brassil, Christopher A Brooks, et al. 2021. ManyClasses 1: Assessing the generalizable effect of immediate feedback versus delayed feedback across many college classes. *Advances in Methods and Practices in Psychological Science* 4, 3 (2021), 25152459211027575.

[8] Xinying Hou, Paulo F Carvalho, and Kenneth R Koedinger. 2021. Drinking our own champagne: Analyzing the impact of learning-by-doing resources in an e-learning course. In *Companion Proceedings of the 11thInternational Conference on Learning Analytics & Knowledge LAK20*.

[9] Kenneth R Koedinger, Paulo F Carvalho, Ran Liu, and Elizabeth A McLaughlin. 2023. An astonishing regularity in student learning rate. *Proceedings of the National Academy of Sciences* 120, 13 (2023), e2221311120.

[10] Kenneth R Koedinger, Jihee Kim, Julianna Zhuxin Jia, Elizabeth A McLaughlin, and Norman L Bier. 2015. Learning is not a spectator sport: Doing is better than watching for learning from a MOOC. In *Proceedings of the second (2015) ACM conference on learning@ scale*. 111–120.

[11] Marsha Lovett, Oded Meyer, and Candace Thille. 2008. JIME-The open learning initiative: Measuring the effectiveness of the OLI statistics course in accelerating student learning. *Journal of Interactive Media in Education* 2008, 1 (2008), 13–13.

[12] Robert C Mathews and Endel Tulving. 1973. Effects of three types of repetition on cued and noncued recall of words. *Journal of Verbal Learning and Verbal Behavior* 12, 6 (1973), 707–721.

[13] Neil W Mulligan and Daniel J Peterson. 2013. The negative repetition effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 39, 5 (2013), 1403.

[14] Kelly Rivers, Erik Harpstead, and Kenneth R Koedinger. 2016. Learning curve analysis for programming: Which concepts do students struggle with?. In *ICER*, Vol. 16. ACM, 143–151.

[15] Nicholas C Soderstrom and Robert A Bjork. 2015. Learning versus performance: An integrative review. *Perspectives on Psychological Science* 10, 2 (2015), 176–199.

[16] John Stamper, Kenneth Koedinger, and Elizabeth Mclaughlin. 2013. A comparison of model selection metrics in datashop. In *Educational Data Mining 2013*.

[17] Masahiro Takimoto. 2012. Assessing the effects of identical task repetition and task-type repetition on learners' recognition and production of second language request downgraders. *Intercultural Pragmatics* 9, 1 (2012), 71–96.