

## POP QUIZZICAL: DOES AUTHORIZING QUESTIONS FOR PEERS IMPROVE LEARNING IN INTRODUCTORY STATISTICS?

Sohee Kang<sup>1</sup> and Justin Slater<sup>2</sup>

<sup>1</sup>Department of Computer and Mathematical Sciences, University of Toronto Scarborough, Canada

<sup>2</sup>Department of Statistical Sciences, University of Toronto, Canada

[sohee.kang@utoronto.ca](mailto:sohee.kang@utoronto.ca)

*Engaging students is critical to fulfilling the learning objectives of any course and is particularly challenging in a remote learning environment. To foster engagement in an introductory statistics course and for weekly participation marks, we employed the online software Quizzical, where students create multiple choice questions based on lecture material that will be answered by their peers. In this paper, we investigate whether the engagement level of Quizzical has any positive association with formal test performance (quiz, midterm, and final) in a large first year introductory statistic course, adjusting for their attitude toward statistics as measured pre-course. We also analyze end-of-term survey data containing qualitative comments on students' perceptions of Quizzical.*

### INTRODUCTION

Many experienced educators will likely agree that crafting a good multiple-choice question is typically more difficult than answering one, and that it requires a deeper understanding of the subject. Using Bloom's revised taxonomy of learning as an example, multiple-choice questions are frequently answered by recalling knowledge or applying a technique; remember and apply are both verbs from the lower half of the taxonomy. In contrast, the highest level of the cognitive domain is necessary to create a question (Krathwohl, 2002).

In recent years, educators have begun to exploit this by creating assessment tasks that not only require students to answer questions, but also to write them, to encourage students' deep learning. Fellenz's (2004) study, which he named Multiple Choice Item Development Assignment (MCIDA) for business students, was an early example (Fellenz, 2004). Students were required to create three sets of three multiple-choice questions (MCQs), each of which contained a stem with one correct answer (the key) and three incorrect responses (distractors). Students were expected to identify the correct answer and explain why it was accurate, as well as explain why the distractors were chosen and why they were incorrect. Students were also asked to determine which level of Bloom's taxonomy would be required to answer the question correctly. Students would receive instructor comments on the quality of their questions between each set of questions, allowing them to improve on their subsequent tries. The best questions were used in the summative MCQ evaluation at the end of the module.

The on-line MCQ authoring, testing, and learning tool called Quizzical (Riggs et al., 2020; Riggs et al., 2014) (<https://quizzical.ca>) was developed to align with Fellenz's (2004) study, and it is fully incorporated with the learning management system at the University of Toronto. Students author MCQ questions and answers with explanations; other members of the class are then able to attempt these questions and leave ratings (see Figure 1.) Another free online tool to facilitate student construction, peer review, and answering of MCQ questions is PeerWise, developed at the University of Auckland (<https://peerwise.cs.auckland.ac.nz>). In PeerWise, students can leave comments about questions and can flag questions they think may be wrong. The system also tracks participation and rewards points and badges for completing tasks, creating a gamification element.

The main difference between the two systems is that Quizzical includes an extra phase in which teaching assistants or instructors can evaluate submitted questions and provide feedback. If the questions cannot be accepted, students can integrate the feedback and resubmit the question, which is closer to Fellenz's (2004) study.

### MOTIVATION

There were several reasons that we employed Quizzical in the introductory statistics course, Statistics I, in the fall of 2021. During the pandemic, the course was taught in a hybrid format where students could sign up for either online or in-person classes. Lecture videos recorded using Zoom in synchronous sessions or direct recordings of lectures were provided to students. We needed to design

course homework that encouraged student learning from the materials and provided regular practice opportunities at the same time. Because procrastination is negatively associated with academic achievement (Kim & Seo, 2015; Michinov et al., 2011), we provided weekly homework practice. Practice retrieving information is also essential for learning new content and storing it in long-term memory (Dunlosky et al., 2013; Karpicke & Blunt, 2011; Roediger III & Butler, 2011). Both formal testing and self-assessment activities, such as online quizzes, were provided to students. Given that authoring original questions requires the highest level of cognition on Bloom's taxonomy (Bloom et al., 1956), we hypothesized that Quizzical is an effective tool for improving learning outcomes and attitudes towards statistics. Another potential benefit is that it encourages weekly problem-solving practice, provides instant feedback, and gives students multiple attempts to demonstrate their understanding without the stress of a formal assessment. Students received a 3% course grade for authoring one question throughout the term and earned a 10% participation mark when they complete at least five practice questions (with at least 60% correctness) pertaining to the previous week's material. This study aimed to explore Quizzical use and its impact on the learning experience and exam results of students enrolled in an introductory statistics course. The overarching research questions for this study are as follows:

1. Does the use of Quizzical increase student engagement in an introductory statistics course?
2. After controlling for baseline attitude toward statistics subjects, which component has the higher impact on students' performance: Quizzical authorship or Quizzical engagement?
3. What are students' perceptions of the benefits and challenges associated with Quizzical?

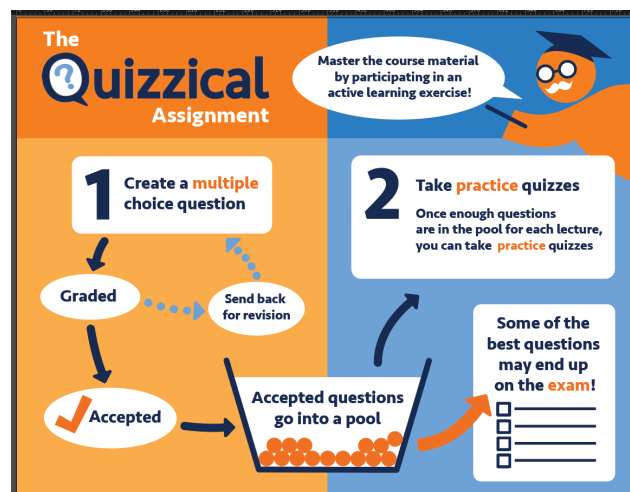


Figure 1. Overview of the Quizzical protocol from <https://quizzical.ca/overview/>

## QUIZZICAL USAGE

At the end of the term enrollment of the Statistics I course, 90% of enrolled students were either first- or second-year students (56% were first-year students and 34% were second year students). Out of 462 students, 303 students provided the consent to participate in this research study. We first examined Quizzical's daily quiz bank usage. Figure 2 shows the average and median attempts per student throughout the term. Spikes in the graph matched with the weekly deadlines of Quizzical participation. The mean number of attempts from each student across the semester was 125.5 questions, and the median was 84 questions.

## REGRESSION MODELING

We extracted some survey questions from the Survey of Attitudes Toward Statistics (SAT-36) (Schau, 1995) to measure students' attitudes towards statistics. We wanted to explore students' general attitude instead of specific attitude factors, so we added 12 selected questions as a measure of students' underlying attitude before they took the statistics courses as a pre-course survey. Based on the median value, we divided Attitude scores into two groups: high and low attitudes. To measure student achievement, we computed a Grade variable as the sum of three variables: average quiz scores (0–15),

midterm scores (0–25), and final exam scores (0–35). Grade is used as the response variable in a regression model for which the explanatory variables are: Quizzical authorship score (0–3); average Quizzical participation score (0–10); and attitude (high and low attitudes). There were 281 students who provided consent and completed the pre-course survey. The mean grade was 48.96 out of 75, and the median was 48.50. Authorship scores have a mean and median of 2.4 and 3 out of 3, respectively, and participation scores have a mean and median of 8.1 and 8.9 out of 10, respectively. The regression model was fit to investigate which contributing factors have a statistically significant effect on the formal assessments of the course.

$$E(\text{Grade}) = \beta_0 + \beta_1 (\text{Quizzical Authorship}) + \beta_2 (\text{Quizzical Participation}) + \beta_3 (\text{Attitude.group}) \quad (1)$$

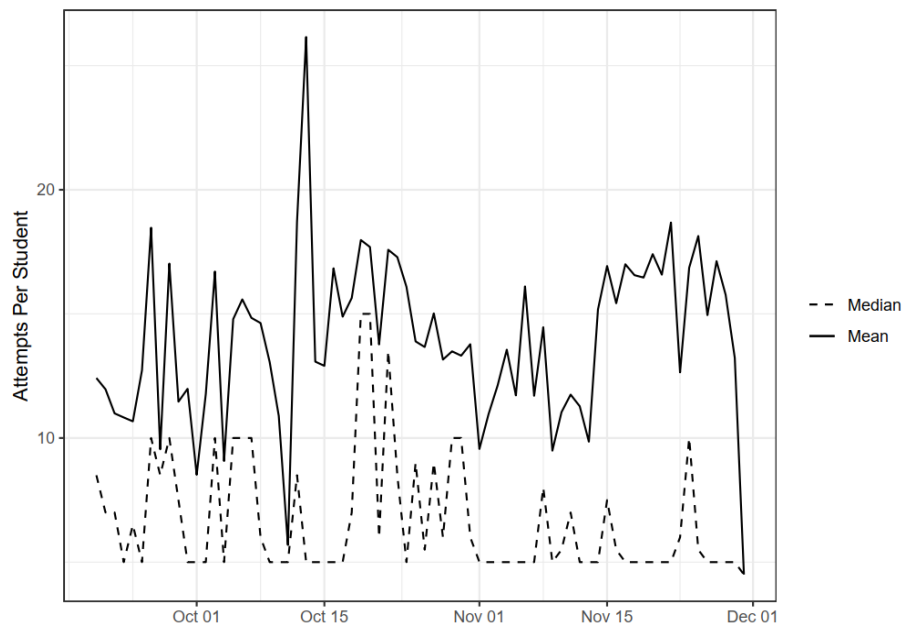


Figure 2. Mean and median number of Quizzical attempts each day

Although too many other factors are missing from the model (based on the low adjusted *R*-squared values), it is noteworthy to observe that Quizzical participation and students’ underlying prior attitude towards statistics are significant factors for grades (see Table 1). After controlling for students’ attitude toward statistics, summative assessment scores are expected to increase by approximately 2.4 when Quizzical participation scores is increased by one unit. The authorship score makes no significant contribution to predicting students’ grades. The weight of the authorship score was quite low (3%), and it was very generously marked. (Most students received full credit.) Furthermore, students only created one question throughout the term, which could explain the lack of association with grade. We also tried including interaction terms in the model such as *Attitude.group* × *Q.Participation* and *Attitude.group* × *Q.Author*, but these were not statistically significant.

Table 1. Regression R output (*n* = 281)

Variable	Estimate	Std.Error	<i>t</i> value	<i>p</i> -value
Intercept	31.77047	2.53655	12.525	< 2e-16 ***
Q.Author	-0.06053	0.72539	-0.083	0.93356
Q.Participation	2.35196	0.32267	7.289	3.27e-12 ***
Attitude.group (Low)	-3.38835	1.27128	-2.665	0.00814 **

Multiple *R*-squared: 0.2071, Adjusted *R*-squared: 0.1986  
*F*-statistic: 24.12 on 3 and 277 *DF*, *p*-value: 6.708e-14

## QUIZZICAL SURVEY

In the final week of class, we sent out a survey to students in order to measure their level of satisfaction with Quizzical. Out of 332 students who participated in the survey, 303 students provided consent for the study. For the first six survey questions, we converted the following categories to a Likert scale of 5 to 1: “Strongly Agree,” “Agree,” “Neutral,” “Disagree,” and “Strongly Disagree.” The description of survey questions and the frequency distribution of responses is presented in Table 2.

Table 2. Quizzical survey questions and frequency distribution tables of responses ( $n = 303$ )

#	Survey Question	Frequency Distribution Table					
1	Creating a question for the assigned lecture week material was helpful to my learning of that particular lecture material.	Scale	1	2	3	4	5
		Count	2	11	41	132	117
2	Practicing questions for the assigned lecture week material was helpful to my learning of that particular lecture material.	Scale	1	2	3	4	5
		Count	0	2	21	108	172
3	Practicing weekly Quizzical questions helped me engage with the course material.	Scale	1	2	3	4	5
		Count	0	6	22	135	139
4	Answering weekly Quizzical questions better prepared me for quizzes and tests in STAB22.	Scale	1	2	3	4	5
		Count	2	16	57	95	133
5	I would like to use peer-generated questions again in future courses.	Scale	1	2	3	4	5
		Count	5	21	76	108	93
6	Creation and use of peer-generated questions created a deeper learning of Statistical concepts which better prepared me for future statistics classes.	Scale	1	2	3	4	5
		Count	3	18	63	144	75
7	What did you think, generally, about the quality of the questions created by your classmates?	Scale	1	2	3	4	5
		Count	2	7	94	163	37

The median score on all questions was 4 ("Agree") or higher. Quizzical's feature that allowed students to practice weekly questions was the most popular among students because it helped them to engage with and learn the lecture content. We also gathered feedback from students in the form of two open-ended short answer questions. We manually read through students' feedback and subjectively categorized the responses into themes and summarize the results in Table 3. Overall, there were more positive comments regarding Quizzical than negative comments. Most of the negative comments are related to the quality of student-generated questions.

## CONCLUSION

In our research, we discovered that weekly Quizzical question practice is a significant predictor of summative course grades after controlling for the underlying attitudes toward statistics. Students' comments on Quizzical complemented our statistical findings indicating they preferred the application's weekly practice component, as opposed to the question authorship component. These results are in line with other studies showing that participating in PeerWise (free online version of Quizzical) has academic benefits for students who write and answer questions (Kay et al., 2020; McQueen et al., 2014; Walsh et al., 2018). There have been also studies that practice assessments have a range of cognitive, metacognitive, and noncognitive benefits (Adesope et al., 2017). Practice assessments can reinforce learning, enhance long-term retention, and can be a strong predictor of exam success (Dunlosky et al., 2013; Karpicke & Roediger III, 2008). However, students in our study completed only one low-grade authoring task, which could explain why authorship scores were not found to be a significant predictor of formal assessment grades.

Table 3. Summaries of Quizzical survey open-ended questions ( $n = 303$ )

Question:	What did you personally find most useful / the best part about using Quizzical?
Theme	Selected Comments
Quizzical encourages regular and low-stress practice. (79 responses)	<p>“...Another great thing about Quizzical is the fact that I knew that if I did it I would get participation marks which not only motivated me to do the work but ensured I got it done every week!”</p> <p>“The review function after every question was a powerful tool I believe because the peer explanations were usually very clear and well written. Overall the quizzical is a very useful method of helping students study. It helps relay topics across to students in a positive manner such that it isn’t stressful allowing them to make mistakes and develop the way we think about questions. While being rewarded. I believe this method of learning is what makes students most successful.”</p>
Quizzical provides immediate feedback with explanations. (21 responses)	<p>“I liked immediately being able to read the justifications for why an answer I selected was incorrect”</p> <p>“... The detailed explanation was very useful as I was able to see where exactly I went wrong when solving the question. I liked using Quizzical questions to study rather than the textbook because the textbook lacks the detailed explanations that Quizzical provides.”</p>
Exam study tool. (21 responses)	<p>“I found Quizzical most useful when I was preparing for quizzes or tests. Before these evaluations I would do 10-20 Quizzical questions to prepare. I enjoyed preparing for the tests this way since it was engaging and the questions tested my knowledge of the material.”</p>
Creating questions was rewarding (10 responses)	<p>“I think it’s interesting to come up with our own questions. Some of the questions were really fascinating, and I enjoyed doing them. Because of the online class this year, I was not able to meet my classmates, but the Quizzical platform made this online class more interactive :)”</p>
Question:	What did you personally find least useful/the worst part about using Quizzical?
Question quality was sometimes poor. (16 responses)	<p>“Many of my colleagues are too perfunctory in asking questions. I think they should work harder.”</p> <p>“I think the accuracy of the question and answers needs to be improved. Often got wrong questions or answers.”</p>
Difficulty was hard to predict. (13 responses)	<p>“The level of difficulty of the questions varied greatly. I would have liked to be able to choose the level of difficulty in order to maximize my learning.”</p> <p>“If there was a particular subject I found really difficult I felt less enticed to do additional questions for fear of losing the 60% threshold.”</p>

One limitation of our study is that we did not validate our extracted 12 survey questions out of the original SAT-36 survey and changed the response scale to 1–5 from 1–7 because our pre-course survey, which was distributed during the first week of the class, included a variety of other questions in addition to attitude. As a result, we did not use the raw scores but rather classified the attitude scores into high and low categories. Our follow-up study should address this problem. Another limitation is that the quality of student-generated questions varied by students’ effort and statistical ability. Due to a shortage of resources for the course, we were unable to allow students’ revisions of their previously submitted questions. Our next step will be to perform a controlled experiment to see if teaching assistant or instructor feedback has a substantial impact on students’ knowledge of the subjects addressed in the questions. Lastly, despite the large sample size, the study was limited to a single course at one school, limiting the findings’ generalizability. There also there may be confounding factors such as a student’s general degree of diligence, although the Quizzical practice component was associated with formal assessment grades.

As statistics educators, we look to develop students into lifelong learners. The initial results presented in this study suggest that the regular practice offered by Quizzical has a positive effect on students' learning outcomes in introductory statistics. In future research, we will look for ways to improve the authorship component so that students can yield greater benefits from online tools such as Quizzical.

## REFERENCES

- Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Education Research*, 87(3), 659–701. <https://doi.org/10.3102/0034654316689306>
- Bloom, B., Englehart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. Longmans, Green.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4–58. <https://doi.org/10.1177/1529100612453266>
- Fellenz, M. R. (2004). Using assessment to support higher level learning: The multiple choice item development assignment. *Assessment & Evaluation in Higher Education*, 29(6), 703–719. <https://doi.org/10.1080/0260293042000227245>
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331(6018), 772–775. <https://doi.org/10.1126/science.1199327>
- Karpicke, J. D., & Roediger III, H. L. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968. <https://doi.org/10.1126/science.1152408>
- Kay, A. E., Hardy, J., & Galloway, R. K. (2020). Student use of PeerWise: a multi-institutional, multidisciplinary evaluation. *British Journal of Educational Technology*, 51(1), 23–35. <https://doi.org/10.1111/bjet.12754>
- Kim, K. R., & Seo, E. H. (2015). The relationship between procrastination and academic performance: A meta-analysis. *Personality & Individual Differences* 82, 26–33. <https://doi.org/10.1016/j.paid.2015.02.038>
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, 41(4), 212–218. [https://doi.org/10.1207/s15430421tip4104\\_2](https://doi.org/10.1207/s15430421tip4104_2)
- McQueen, H. A., Shields, C., Finnegan, D. J., Higham, J., & Simmen, M. W. (2014). PeerWise provides significant academic benefits to biological science students across diverse learning tasks, but with minimal instructor intervention. *Biochemistry and Molecular Biology Education*, 2(5), 371–381. <https://doi.org/10.1002/bmb.20806>
- Michinov, N., Brunot, S., Le Bohec, O., Juhel, J., & Delaval, M. (2011). Procrastination, participation, and performance in online learning environments. *Computers & Education*, 56(1), 243–252. <https://doi.org/10.1016/j.compedu.2010.07.025>
- Schau, C., Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (1995). The development and validation of the survey of attitudes toward statistics. *Educational and Psychological Measurement*, 55(5), 868–875. <https://doi.org/10.1177/0013164495055005022>
- Riggs, C. D., Kang, S., & Rennie, O. (2020). Positive impact of multiple-choice question authoring and regular quiz participation on student learning. *CBE-Life Sciences Education*, 19(2). <https://doi.org/10.1187/cbe.19-09-0189>
- Riggs, C. D., Sutherland, B., Irani, A., & Patterson, J. (2014). “Quizzical”: A student-authored, online, multiple choice question writing and learning assessment tool. In *Proceedings of the ICED 2014 Educational Development in a Changing World*. International Consortium for Educational Development. [http://www.iced2014.se/proceedings/1673\\_Riggs.pdf](http://www.iced2014.se/proceedings/1673_Riggs.pdf)
- Roediger III, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Science* 15(1), 20–27. <https://doi.org/10.1016/j.tics.2010.09.003>
- Walsh, J. L., Harris, B. H., Denny, P., & Smith, P. (2018). Formative student-authored question bank: Perceptions, question quality and association with summative performance. *Postgraduate Medical Journal*, 94(1108), 97–103. <https://doi.org/10.1136/postgradmedj-2017-135018>