Exploring the Impact of Instructor Questions in Community College Algebra Classrooms

Linda Leckrone Amin Ullah Amy Foster-Gimbel

Dmetri Culkar Holly Whitney Annemarie McDonald Vilma Mesa

University of Michigan

We describe a process to characterize the questions asked by instructors and students in community college algebra courses. The goal is to measure the quality of mathematical questions that can speak to the level of student cognitive engagement with mathematics and to connect that quality with student outcomes in the course. As a first step, we explore the relation between frequency of different types of questions and other variables collected in the project. We seek to engage the audience in discussing the affordances and limitations of this work for assessing quality of instruction in connection to students' performance.

Keywords: Community Colleges, Algebra, Question Quality, Student Outcomes

Questions are a form of discourse that have the potential to open up a conversation (Martin & White, 2005). Questioning in mathematics classrooms can play a significant role in advancing student engagement with mathematical content. In community college classrooms, in which the predominant mode of instruction is lecture, instructors say they use questions as a tool to keep students engaged with the content (Burn & Mesa, 2017). While most research has documented that questioning is an important classroom practice, it is unclear how questioning is correlated with student outcomes. Most of the literature on questions documents what instructors and students do when questioning takes place; describing for example the frequency of certain types of questions (Paoletti et al., 2018) or the types of reasoning that they may elicit (Temple & Doerr, 2012). However, assessing the impact of different types of questions has not been pursued, mainly because the work of analyzing classroom discourse is time-consuming, and is typically done on a small-scale basis with few instructors and lessons.

We focus on community college algebra courses because their high failure rate is seen as a reason for students abandoning their plans to complete a degree (Bahr, 2010). As part of a large-scale study of algebra instruction at community colleges, we sought to establish whether and how, the quality of questions relates to various student outcomes in the course. In this preliminary report, we focus on the process of developing a system to code the quality of questions asked by community college instructors teaching algebra courses, and a preliminary analysis that seeks to link the types of questions instructors ask in the classroom with student outcomes in those courses. Our focus was to accurately and reliably code questions asked during instruction and use frequencies of those codes to explore relationships to student performance. Because questions play a predominant role in community college mathematics classrooms, if there is a connection between the quality of questions and student performance, then improving how questions are used in the classroom may lead to more opportunities for student learning. If questioning practices do indeed have an impact on student outcomes in these courses, one could envision a way to use questioning as leverage for improving instruction in ways that can have a real impact on students.

Supporting Literature

Cognitive theory provides strong support for engaging learners in activities that encourage them to draw on their knowledge (factual, procedural, conceptual, metacognitive) using an array of cognitive processes (e.g., remember, apply, evaluate, etc., Anderson et al., 2001). In an environment in which lecture dominates, questions can open a space for cognitive engagement (Mesa & Chang, 2010). The literature suggests that community college mathematics instructors ask a large number of questions as they teach (Mesa, 2010; Mesa & Lande, 2014). These, and other studies of lecturers and faculty, also suggests that that instructors tend to ask questions that for the most part require students to recall information they already know; questions that demand higher level reasoning are asked less frequently (Larson & Lovelace, 2013; Mesa, Celis, & Lande, 2014; Paoletti et al., 2018). However, the quality of questions that instructors ask can encourage students' critical thinking (Boerst, Sleep, Ball, & Bass, 2011). Questions that require students to go beyond what the instructor has presented in the lesson may compel students to bring in information or make connections beyond what is known in the class. Such questions have higher levels of cognitive demand than questions that ask students to recall facts they are expected to know. We may then expect that students in courses whose instructors ask them questions that challenge their thinking or that demand high cognitive work will have better performance.

Methods

As part of the larger project, in fall 2017, 40 different instructors were video recorded teaching at least two lessons in intermediate and college algebra classrooms on one of three topics: linear, rational, or exponential equations and functions. The instructors, who taught at six different community colleges in three different states, volunteered to take part in the study. They filled out questionnaires on beliefs, personal information, and a test of their Mathematical Knowledge for Teaching Algebra (MKT-A). Their students also filled out various questionnaires, including a test measuring covariational reasoning, which was administered twice in the semester: two weeks after the beginning of the term and two weeks before the end of the term. For developing ways to capture the quality of instructor questioning we coded 37 lessons from 15 instructors selected randomly from the high, average, and low scores in the MKT-A questionnaire. Instructor and student scores were not shared with the coders.

To code the questions in a lesson we adapted the taxonomy proposed in Mesa and Lande (2014), which includes two major categories of questions, mathematical and non-mathematical. Mathematical questions were coded as realized if the students provided an answer or the instructor waited a sufficient amount of time after each question (5 seconds or more), and unrealized if there was no answer and little or no wait time. Mathematical questions were placed into three categories: authentic, quasi-authentic, and inauthentic. In this paper, we focus only on mathematical questions. Table 1 presents the categories of questions of the coding process, their definitions, and examples.

Table 1: Coding system for mathematical questions in videos of community college algebra lessons.

Realized Authentic (RA) Questions that, if answered, would require students to use information beyond what they have learned in class. These are often open-ended questions.	T: So, did you guys come up with an example of a situation where the input, right, how I evaluate the output, changes based off the input? What are you coming up with? S: We decided that money could be {... }

Realized Quasi-authentic (RQ) Questions that require some knowledge and have limited possible answers. Questions that can be answered with material to be recalled that is new in that lesson.	T: Anybody see how you could clear all these denominators with a valid mathematical process? S: Inaudible response
Realized Inauthentic (RI) Questions that require using information that is known, expected to be known	T: But remember, $x + x$ is actually what? S: $2x$
Unrealized (UA, UQ, UI) Questions that are not answered by students and the instructor waits less than five seconds for a response	T: And speaking of defining terms, what the heck is a rational function? (no pause) Well, it's {answers question}

Procedure

Seven coders were recruited to assist in the coding of lessons. One difficulty of working in a large group of people, when decisions need to be reached, is the threat of groupthink. Groupthink is "a concurrence-seeking tendency that can impede collective decision-making processes and lead to poor decisions" (Choi & Kim, 1999).The most effective ways to combat groupthink is with the encouragement of dissenters. In our group, we created an environment where all members felt comfortable voicing their opinion. Disagreement among coders was not seen as a hindrance to the study or an obstacle to be overcome, but rather as a key aspect of the objectivity and reliability of our process. By emphasizing each individual's viewpoint, we ensured that the final coding was never reached solely through a member's concession and that every code given had strong collective support. To do the coding, each video file was given to one coder who transcribed and coded the questions. Then, the codes were hidden, and the video file and transcription were given to a second coder, who watched the video again to find any questions the first coder may have missed and to code the questions. The coders resolved any discrepancies by consensus; we held weekly meetings to refine the coding process using understandings from the resolution of discrepancies. To measure inter-rater agreement, we calculated either Cohen's κ (2 coders) or Fleiss' κ (3 coders or more) prior to resolving the discrepancies. Initially κ values were low (around 0.3) but they improved as the protocol was refined after discussions (~0.6-0.8).

There were two areas of difficulty: (1) differentiating between rhetorical questions and unrealized authentic questions, and (2) deciding the authenticity based on the knowledge available to the class. For (1), when introducing a new topic an instructor may say: "What the heck is a rational function?" If students did not respond to the question, and the instructor did not leave time for students to respond, some coders code this question rhetorical and some as authentic unrealized. The final decision was to code these questions as unrealized mathematical acknowledges that they could prompt discussion of new topics and therefore help students create new connections, were the instructor interested in giving students opportunity to engage with the questions. For (2), because instructors in different community colleges introduce topics in different orders, material that may be completely new in one class might have been covered in a previous lesson in another. This made it difficult to know whether a question related to new material or students' prior knowledge. In order to manage this issue, we used (a) the course syllabi to resolve disagreements about what constitutes new material and prior knowledge, (b) relied on coders with experience teaching the course to decide, or (c), when neither (a) or (b) was conclusive, used the "generous" coding approach and assigning the higher level. These decisions helped improve the inter-rater agreement.

Preliminary Findings

We coded 8,323 instructor questions of which about 20% were either authentic or quasi-authentic (n = 1,600) and 40% (n = 3,332) were inauthentic. The average wait time across all 37 lessons after a question was asked was about one second (0.99); only 4% of the questions had a wait time of 5 seconds or more. Across all lessons, instructors asked, on average, five questions every two minutes (2.49 questions per minute). These averages mask variations by lessons. Because the lessons have different lengths, we calculated the rate of questions (# of questions per lesson divided by lesson duration) to facilitate cross-comparison.[1] Table 2 presents descriptive information for several variables that were created.

Table 2: Mean and Standard Deviation for Variables in the Study. (N=37)

	Mean	SD
Rate per minute of Realized[a] Authentic and Quasi Authentic Questions[b]	.116	.069
Rate per minute of Unrealized Authentic and Quasi Authentic Questions[b]	.083	.049
Rate per minute of all Authentic and Quasi-authentic Questions[b]	.498	.402
Rate per minute of Inauthentic Questions[b]	.982	.528
Proportion of Authentic and Quasi-authentic Questions[c]	.198	.091
Average Wait Time (in seconds)	.986	.809
Proportion of Questions for which there is a 5s or more Pause	.048	.047
Normalized Gain in Scores on Test of Knowledge[d]	.102	.064
Final Grade[e]	.589	.153
Proportion of Students Passing the Course	.755	.192
MKT-A[f]	27.6	2.983

Notes: a. A question is realized when there is a student response or a pause of five seconds or more after a question has been posed. b. Estimated as sum of the two categories of questions and divided by the duration of the lesson. c. Estimated as the sum of the two categories and divided by the total number of questions asked. d. Difference between end of term and beginning of term scores divided by the number of questions not answered in the beginning of term test. e. Average final grade on a scale of 100, divided by 100. f. Instructor score in the MKT-A; maximum points in the test was 34.

In these lessons, the instructors asked about one inauthentic question per minute, and about one authentic or quasi-authentic question every two minutes. On average the gains in the student test of knowledge were small: about one and a half more questions responded correctly and about 10% of gain at the end of the term. The average grade in the courses was 59%, and about 76% of students passed their course. To explore possible relations between the quality of questions and the student outcomes, we tested correlations between these variables, using a non-parametric test (Kendall's τ) as the distributions of these variables were not normal.

---

[1] We also averaged these rates to obtain a single value for each instructor, under the assumption that while there might be differences in lessons that may result in more or less use of different types of questions, the way in which instructors use questioning is a feature of their instruction and not as dependent on the content at stake. In this proposal, we only present analyses at the lesson level.

We found positive and statistically significant correlations between the rate of realized authentic and quasi-authentic questions and the MKT-A score ($\tau = .207, p < .05$) and the percentage of questions for which there is a pause of 5s or more ($\tau =.231, p < .05$). We also found negative and statistically significant correlations between the rate of inauthentic questions and the proportion of students who pass the course ($\tau = -.209, p < .05$) and the MKT-A score ($\tau = -.205, p < .05$). Finally, there was a positive and statistically significant correlation between the proportion of authentic and quasi-authentic questions and the MKT-A score. This suggests that instructors with higher MKT-A scores tended to ask (1) a higher rate per minute of authentic and quasi-authentic questions for which they wanted an answer (i.e., questions that were realized) and (2) a lower rate of inauthentic questions per minute. The rate of questions per minute for authentic and/or quasi-authentic questions was positively correlated with the percentage of questions for which there was a pause. We anticipated this result. And as a consequence of the coding, we believe this suggests that when instructors paused it was likely that they did so for an authentic or quasi-authentic question. We found a positive and borderline statistically significant correlation between the average length of the pauses after the question and the percent grade ($\tau = .192, p = .053$) and the proportion of students passing the course ($\tau = .19, p = .056$). Thus, as the duration of the wait time increased, the pass rate in the course increased (or in courses in which more students passed, there was more wait time after questions posed).

Discussion

These are preliminary analyses, but they suggest, first, that it might be detrimental to ask too many inauthentic questions per minute, and second, that instructors with more knowledge of algebra for teaching (as measured with the MKT-A test) will pose more authentic and quasi-authentic questions with longer pauses. These findings also suggest that when instructors ask authentic and quasi-authentic questions without giving students opportunity to respond or time to think about the questions, such decision might be detrimental for student outcomes. While using rhetorical questions to introduce a topic can be a strategy to capture students' interest and attention in the end they might not be as effective. The connection between scores in the MKT-A and use of types of questions is promising, as it suggests concurrent validity between the two measures. The lack of relationship between the types of questions and student performance on the test of knowledge might be related to the generic character of the coding system.

Questions for Discussion

1. We coded mathematical questions using six categories, based on the level of cognitive demand required to answer it and whether or not the question was realized. What are other possible ways of measuring the quality of questions?
2. Interpreting the anticipated cognitive demand of a question is challenging. What measures should be included to increase the reliability of this type of coding?

Acknowledgments

Funding for this work was provided by the National Science Foundation award EHR #1561436. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We also thank Cody Michael and the faculty who participated in the project. Without them this work would not have been possible.

References

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., . . . Wittrock, M. C. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing*. New York, NY: Longman.

Bahr, P. R. (2010). Revisiting the efficacy of postsecondary remediation: The moderating effects of depth/breadth of deficiency. *Review of Higher Education, 33*, 177-205.

Boerst, T. A., Sleep, L., Ball, D. L., & Bass, H. (2011). Preparing teachers to lead mathematics discussions. *Teachers College Record, 113*(12), 2844-2877.

Burn, H., & Mesa, V. (2017). Not your grandma's lecture: Interactive lecture in the CSPCC two-year cases. *MathAmatyc Educator, 8*(3), 24-29.

Choi, J. N., & Kim, M. U. (1999). The organizational application of groupthink and its limitations in organizations. *Journal of Applied Psychology, 84*(2), 297-306. doi:10.1037/0021-9010.84.2.297

Larson, L. R., & Lovelace, M. D. (2013). Evaluating the efficacy of questioning strategies in lecture-based classroom environments: Are we asking the right questions? *Journal on Excellence in College Teaching, 24*(1), 105-122.

Martin, J. R., & White, P. R. R. (2005). *The language of evaluation: Appraisal in English*. New York: Palgrave-Macmillan.

Mesa, V. (2010). Student participation in mathematics lessons taught by seven successful community college instructors. *Adults Learning Mathematics, 5*, 64-88.

Mesa, V., Celis, S., & Lande, E. (2014). Teaching approaches of community college mathematics faculty: Do they relate to classroom practices? *American Educational Research Journal, 51*, 117-151. doi:10.3102/0002831213505759

Mesa, V., & Chang, P. (2010). The language of engagement in two highly interactive undergraduate mathematics classrooms. *Linguistics and Education, 21*, 83-100.

Mesa, V., & Lande, E. (2014). Methodological considerations in the analysis of classroom interaction in community college trigonometry. In Y. Li, E. A. Silver, & S. Li (Eds.), *Transforming math instruction: Multiple approaches and practices* (pp. 475-500). Dordrecth, The Netherlands: Springer.

Paoletti, T., Krupnik, V., Papadopoulos, D., Olsen, J., Fukawa-Connelly, T., & Weber, K. (2018). Teacher questioning and invitations to participate in advanced mathematics lectures. *Educational Studies in Mathematics, 98*(1), 1-17.

Temple, C., & Doerr, H. (2012). Developing fluency in the mathematical register through conversation in a tenth-grade classroom. *Educational Studies in Mathematics, 81*, 287-306.