Characteristics of Questions That Promote Rich Mathematical Discussions

Holly Zullo, *Carroll College*
Jean McGivney-Burelle, *University of Hartford*
Kelly Cline, *Carroll College*
Ann Stewart, *Hood College*
Chris Storm, *Adelphi University*
David A. Huckaby, *Angelo State University*
Tim Melvin, *Carroll College*

Introduction

Classroom voting is a form of pedagogy in which the instructor poses multiple-choice questions and students respond using colored cards or student-response systems, also known as clickers, to indicate their answers. [Note: There are over 2,000 multiple-choice questions designed for classroom voting in college mathematics courses available at http://mathquest.carroll.edu.] A significant benefit of classroom voting in mathematics is that all students can engage with the content as they consider a question, discuss it with their peers in small groups, and cast their vote. After the vote, the instructor facilitates a class-wide discussion of the question. Both students and instructors get immediate feedback on how students are making sense of the material, which can then inform the teacher's instructional decisions and prompt students to reflect on and perhaps modify their understandings. This article focuses on these class-wide, postvote discussions.

Our position is that a distinguishing feature of good questions is that they promote lively class-wide discussions in which several students participate and clearly articulate their thinking and in which students either challenge or support each other's thinking. While there are anecdotal data lending support for assertions of the characteristics of good questions, there have been no studies that examine the quality of class-wide discussions following voting questions and how these data might point towards new insight into the ingredients for good questions. The purpose of this article is to present data we have collected on the characteristics of questions that promoted highly rated discussions across a range of college mathematics courses at a variety of two- and four-year institutions. In addition, we present data on the intent of an instructor in asking a question and examine how this intent relates to the potential for a good discussion.

Data Collection and Analysis

Data collection for this project occurred between fall 2010 and spring 2012 and involved six different instructors, at five institutions, teaching 17 classes. The course topics included liberal arts mathematics, precalculus, calculus, multivariable calculus, linear algebra, and differential equations. In each class, the instructor used classroom voting on a regular basis. In total, there were 1345 voting events recorded with postvote discussions that were rated by either the instructor or by a student or other faculty observer who was specifically asked to rate the discussions. When the instructor did her/his own rating, it was either done as class was happening, or the rating was done later by reviewing a video or audio recording of the class period.

Each discussion was rated along four dimensions that we judged to be important markers of a good discussion. These dimensions are *participation*, *articulation*, *energy level*, and *peer response*. Raters were asked to score each dimension on a scale from 0 to 2. A cumulative score is obtained by summing the scores across all four dimensions. (See Table 1.)

High ratings in participation arose either from many students volunteering or from an instructor sensing unresolved issues and, therefore, deliberately prolonging a discussion by calling on more students. In evaluating articulation, discussions were rated high on this dimension when students communicated coherent thoughts and made good use of terminology, regardless of the accuracy of their answers. A high energy level score indicates student enthusiasm well beyond the norm for the class. Finally, the best ratings in the peer response category were given when students were primarily addressing each other and each other's reasoning as opposed to presenting their own reasoning in isolation.

Since different people performed the discussion ratings and we were unable to establish inter-rater reliability, we do not compare ratings across courses. Another issue with cross-course comparisons is the difference between classroom

	High (2)	Medium (1)	Low (0)
Participation	Many students involved	A few irregulars	Just the regulars
Articulation	Most students use terms well and have clear arguments	Some of each or mostly "OK" articulation	Most students use vague terms, unclear arguments
Energy Level	Discussion prolonged by student enthusiasm	Most students alert and engaged, some energetic	Zzzzz… students ready to move on
Peer Response	Mostly student–student interactions	Some student–student interactions	Mostly student–teacher interactions

Table 1: Discussion Rating Rubric

cultures. For example, some classes are naturally energetic, leading to high ratings on energy level for almost every question, while others (even when taught by the same instructor) are naturally very quiet and rarely produce a high rating on energy level. Therefore, we examine discussion ratings on a class-by-class basis. Even though the maximum possible score is eight, in some classes, due either to the nature of the class or the disposition of the rater, the highest score was only six or seven. Also, there were vastly different numbers of rated discussions from the various courses, ranging from a low of 13 to a high of 200. In the end, we decided to determine the top discussion rating for each course and then examine all questions from that course that received that rating. This yielded a collection of 29 questions, or 1 to 2 per course. In the next section, we examine some of these questions and some of the characteristics that give them the potential to support a strong classroom discussion. Later, we introduce and incorporate the instructor's intent in asking a question into the analysis.

Results

By identifying the main themes and characteristics of questions that generate good discussions, we hope to improve question writing as well as question selection from existing databases. We have studied the set of 29 questions we identified as being the best discussion generators, and we have identified five main characteristics, with many questions containing several of these characteristics. These five categories are shown in Table 2, along with the number of the top questions that fell into each category. We will discuss the precise meaning of and give examples of questions in each of these categories.

Question category	Number of 29 questions in each category
Informal language and personal experience	7
Application	4
Visual	8
No numbers and general cases	10
Difficult concepts and multi-step problems	11

Table 2: Number of questions displaying each characteristic of good discussion generators

Déjà Vu, Almost (Informal language and personal experience)

Many of the questions that generated the highest-rated discussions used natural and informal language, rather than mathematically precise language. When students are not given the necessary mathematical vocabulary in the question, they must select the appropriate vocabulary themselves as they try to give careful explanations of their answers, often based on their personal experiences. The lack of mathematical vocabulary also often means that students cannot rely on this as a key for determining an approach to solving a problem; they may instead rely more on intuition and may use a wider variety of approaches. An example of this is the rain drop problem from vector calculus, shown in Figure 1, which was presented and discussed in Terrell (2011).

> Assume that it is raining and the droplets all fall straight down with the same speed. You want to minimize how wet you get.
>
> (a) It is better to walk through the rain.
>
> (b) It is better to run through the rain.
>
> (c) It doesn't matter if you walk or run, you will get just as wet either way.

Figure 1: Example of a question posed with informal language

This question requires students to rely on their intuition based on personal experience. Ideally, students tap into their informal experience walking in the rain and start to make connections with more formal mathematical ideas. Eventually, perhaps after some prodding by the instructor, students work their way to talking about flux through various planes.

Good discussion generators are also frequently built around words that have very precise meanings in mathematics but are used without mathematical precision in everyday language. For instance, in mathematics the if-then statement has a very precise logical structure, but common usage often fails to duplicate this structure. An example of this type of question, written by the Cornell Good Questions project (NSF CCLI DUE-0231154), is shown in Figure 2.

> Your mother says, "If you eat your dinner, you can have dessert." You know this means, "If you don't eat your dinner, you cannot have dessert." Your calculus teacher says, "If f is differentiable at x, f is continuous at x." You know this means
>
> (a) If f is not continuous at x, f is not differentiable at x.
>
> (b) If f is not differentiable at x, f is not continuous at x.
>
> (c) Knowing f is not continuous at x does not give us enough information to deduce anything about whether the derivative of f exists at x.

Figure 2: Example of a question where the language helps generate discussion

This question deals directly with an instance where common usage of if-then differs from the precise mathematical use, and students take many approaches in their answers. The "standard" answer is to do what students do best in mathematics classes: look at the example and mimic the structure of it, thus arriving at answer (b). Some students have had calculus before and may remember the relationship between continuity and differentiability and, hence, they arrive at answer (a). Often, though, these students have lingering questions about the mother's statement. Other students recognize that the mother's logic does not match mathematical logic, and they reason their way to the correct logical equivalence.

One important marker in becoming a mature mathematician is appreciating the precision of mathematical language and the importance of mathematical definitions. Questions that explore the interaction between colloquial language and mathematical language, especially those that also tap into students' experiences, can serve a very important role in a student's development and provide a solid platform for good classroom discussions.

How am I ever going to use this? (Application)

Applied scenarios and questions requiring physical interpretation give students an opportunity to translate mathematics into a different context and can open the door for very fruitful discussions. Often, these questions are phrased in natural language, as described previously. However, even when phrased mathematically, these questions can raise significant issues for discussion. An example is given in Figure 3.

> For which of the following predator–prey population models is the predator most successful at catching prey?
>
> (a) $\dfrac{dx}{dt} = 2x - 3xy; \quad \dfrac{dy}{dt} = -y + \dfrac{1}{2}xy$
>
> (b) $\dfrac{dx}{dt} = x(1-4y); \quad \dfrac{dy}{dt} = y(-2+3x)$
>
> (c) $\dfrac{dx}{dt} = x(3-2y); \quad \dfrac{dy}{dt} = y(-1+x)$
>
> (d) $\dfrac{dx}{dt} = 4x\left(\dfrac{1}{2} - y\right); \quad \dfrac{dy}{dt} = 2y\left(-\dfrac{1}{2} + x\right)$

Figure 3: Example of a question using an applied setting

This question involving predator–prey models has been effective in generating deep discussion. Students tend to focus on the coefficients of x in the $\dfrac{dx}{dt}$ equation and of y in the $\dfrac{dy}{dt}$ equation, rather than on the coefficients of the interaction terms.
Thus, this question provides students with an opportunity to fully analyze the meaning of the coefficients in the predator–prey differential equations model. It lays the groundwork for a good discussion as students explore the connection between the numbers in the equation and the consequences of those numbers in the behavior being modeled.

I just don't see it! (Visual)

Students tend to have significant trouble with problems involving graphs, geometric interpretation, and visualization. Over one-fourth of our top-rated questions involve some sort of visualization. We give an example in Figure 4.

> Which of the following integrals is equal to $\int_0^3\int_0^2\int_0^y f(x,y,z)\,dz\,dy\,dx$?
>
> (a) $\int_0^2\int_0^3\int_0^y f(x,y,z)\,dz\,dx\,dy$
>
> (b) $\int_0^2\int_0^3\int_0^y f(x,y,z)\,dz\,dy\,dx$
>
> (c) $\int_0^3\int_0^2\int_0^y f(x,y,z)\,dx\,dy\,dz$
>
> (d) $\int_0^3\int_0^2\int_0^z f(x,y,z)\,dy\,dz\,dx$

Figure 4: A question requiring visualization

Students who sketch the region of integration typically do not have difficulty with this question. However, many students struggle with that visualization and try to rely on memorizing a process for changing the limits of integration. The popular answers to this question are (a) (correct) and (d). A good discussion here elicits information not only about why answer (a) is correct, but also why answer (d) is incorrect, probably presenting the region of integration in (d) and how it differs from the given region. Much as application questions require students to move fluently between mathematics and the real world, visualization questions often require students to move between algebraic information and geometric information.

I thought math was about numbers! (No numbers and general cases)

To break down a problem that doesn't have any numbers, students must be able to conjure up their own numbers to test or reason based upon algebra, or axioms and prior theorems, which can lead to a wide variety of solution approaches and hence an interesting discussion. We present two such questions here. Figure 5 gives an example from a liberal arts mathematics course, and Figure 6 shows an example from a linear algebra course.

> Which is correct?
>
> (a) If you have a majority of the votes then you have a plurality.
>
> (b) If you have a plurality then you have a majority of the votes.
>
> (c) Both statements are false.

Figure 5: Example of question without numbers from a liberal arts mathematics course

Students have to consider the meanings of the terms majority and plurality and then understand them in relationship to one another. Often, students rely on specific cases to justify their answers. For example, students who reason correctly will make statements such as, "If a candidate receives more than 50% of the votes, then they will have more than everyone else." They may also add, "But if there is an election with more than two candidates, then one candidate can have 40% of the votes while each of the other candidates has 30% of the votes and the winner will have a plurality but not a majority of the votes." When this question was posed in class an interesting discussion ensued when a student who correctly voted for (a) then went on to advocate for accepting (b) as an answer in the case of a two-candidate election. This student's response suggests this student is transitioning from an immature view of mathematics to a more sophisticated understanding of mathematical abstraction and universal statements. Students debated whether the way the question was posed meant that we must assume that the statement is true for any case.

This category of questions particularly lends itself to problems with solutions that are supported by multiple ways of thinking. In Figure 6, we see a question that asks about the number of solutions to $Ax = b$ where neither A nor b is given. In the given form, the question is easy to answer if the student has already made a connection between the rank of A and the number of solutions to $Ax = b$, but otherwise the student must generate a sample matrix to test or reason through the connection between rank and the number of solutions. When this question was used in class, the first student to speak gave a perfect answer, but, as is often the case, the other students failed to recognize that. The students with incorrect answers jumped in, and a great class discussion resulted.

> Suppose a 4×4 matrix A has rank 4. How many solutions does the system $Ax = b$ have?
>
> (a) 0
>
> (b) 1
>
> (c) Infinite
>
> (d) Not enough information is given.

Figure 6: Example of question without numbers from a linear algebra course

More than meets the eye! (Difficult concepts and multistep problems)

Good discussion generators do not always jump out at us as faculty. Often, it is sufficient to have a question on a traditionally difficult topic, even if the question itself is somewhat bland. Along the same lines, students always have more trouble with multiple-step problems, and in the case of voting and discussions, these questions provide an opportunity to discuss each of the steps involved. In both of these cases, the questions should have very carefully constructed distracters to catch a variety of common mistakes. Eleven of the 29 questions identified as good discussion generators fell into this category, making this the most common grouping.

The problem shown in Figure 7 is an example of a question on a difficult topic. Students who confuse the notation for function composition with multiplication would incorrectly vote for (b). If students fail to complete both parts of the substitution, they will incorrectly answer (d), or if they perform the function composition in the wrong order, they will incorrectly answer (a). Given that this seemingly straightforward question is rather complex and includes common mistakes as

Given $f(x) = x + 1$ and $g(x) = 3x^2 - 2x$, what is the composition $g(f(x))$?

(a) $3x^2 - 2x + 1$

(b) $(3x^2 - 2x)(x + 1)$

(c) $3x^2 + 4x + 1$

(d) $3(x + 1)^2 - 2x$

Figure 7: Example of question focused on a difficult topic

distracters in the answers, this problem can produce interesting class discussions.

The integration question already presented in Figure 4 also fits into this category. Students frequently try to do this problem in a single step, moving directly from the integral given to the new ordering. Instead, this problem really requires two steps: (1) determining the region of integration and (2) writing the limits for the new order of integration. Each of these steps, and the very necessity of two steps, provides opportunity for discussion.

In summary, the key to all of our top-rated discussion questions is that they include one or more of the following features:

- Common language that taps into students' personal experiences
- Applications of mathematical ideas
- Graphs, geometrical representations, or visualization
- Investigations of general cases or problems without numbers
- Emphasizing difficult concepts or multistep problems

Further, in most of these questions or problems there are multiple answers that are plausible to students, or there are multiple paths (maybe all correct and maybe not all correct) to an answer, and the question focuses on a topic of some difficulty.

Instructor Intent

Perhaps of equal importance to what question is being asked in the classroom is the context in which that question is asked and why the instructor is asking it. Classroom voting can be used in many ways, and each instructor categorized each question according to their intent in using the question: practice, introduction, deep/probing, and review. A question that might be quite straightforward when used for practice can suddenly become very interesting when used before the students have received formal instruction on the topic. Similarly, some questions were classified as "deep" because of how the question was used, not because it is inherently deep. For example, if a topic has been taught from an entirely algebraic perspective and a question asking about graphical interpretation is then posed, this question would probably be put in this category, whereas it might be considered a practice question if the class had already explored the graphical perspective.

Table 3 shows the percentages of questions in each of these categories for two different pools of questions: all of the 1345 (nondistinct) questions whose discussions we rated and the 29 (distinct) questions that produced the highest-quality discussions. We see that a disproportionate number of deep/probing questions generated strong discussions, which is what we would expect. Similarly, questions used for practice can still generate good discussions, but they are underrepresented in the group of top discussion generators.

Instructor Intent	Percent of All Questions	Percent of Top-Rated Questions
Practice	58	31
Introduction	16	10
Deep/Probing	13	55
Review	13	3

Table 3: Distribution of questions by instructor intent

Given the fixed amount of time we have with students in the classroom, this data suggests we should seriously consider questions intended to deepen student understanding if our goal is to promote good classroom discussions about mathematics. An example of a question that gives an instructor the opportunity to use it either for practice or to deepen understanding is given in Figure 8.

Does the harmonic oscillator described below oscillate?

$$3\frac{d^2y}{dt^2} + 4\frac{dy}{dt} + y = 0$$

(a) Yes

(b) No

Figure 8: A question that can be used for practice or deepening

At the point when the question in Figure 8 was used, students knew how to solve this type of second order differential equation, but the qualitative interpretation of the solutions had not yet been discussed. This question was used to deepen student understanding by getting them to explore how the solution can be used to analyze the qualitative behavior of the equation. In addition, since the question did not explicitly ask for the solution to the differential equation, the students had to recognize that taking steps towards finding the solution might be useful. Thus, this question has characteristics from more than one of the categories that we identified (i.e., visual and multiple step) and is being used with the intent of deepening understanding. While all of these might be easy steps for a trained mathematician, it became clear in the discussion that these were big leaps for the students, and a lively discussion ensued.

Top-rated Discussion Questions and Student Accuracy in Voting for the Correct Answer

Looking just at the top-rated discussion questions, the mean percentage of students voting for the correct answer is 41%, and the median percentage is 43%. These numbers are quite a bit lower than might usually be expected. For example, Cline, Zullo, and VonEpps (2012) found that the mean percentage of students voting correctly on differential calculus questions was 67%. This supports the idea that we can use past voting results to identify some of the more effective discussion questions.

Conclusion

In this article, we presented a scheme for identifying key characteristics of mathematics classroom voting questions which led to class-wide discussions that were highly rated by instructors or outside observers. By examining the questions that led to the most highly rated discussions, we found common themes of (1) natural and/or colloquial language drawing on student experiences, (2) applied settings, (3) requiring visualization, (4) a lack of numbers, and (5) difficult concepts and multistep problems. These categories are intended to help mathematics instructors interested in facilitating meaningful classroom discussions to write better questions. Further, awareness of these categories may be useful to instructors who are selecting a few questions from a large collection to use in a particular lesson. In addition, we have examined the role that instructor intent and timing in posing a question can play in achieving a good classroom discussion on a voting question. A majority of the top-rated discussions were produced when the instructors used questions to deepen student understanding by asking them to work with a concept in a new way, rather than asking students to practice a technique that already had been explicitly taught.

Acknowledgments

This paper is based on work supported by National Science Foundation Grant DUE 0836775. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Cline, K., Zullo, H., & VonEpps, L. (2012). Classroom Voting Patterns in Differential Calculus. *PRIMUS*, *22*(1), 43–59.

Terrell, M. (2011). Good questions for mathematical education: An example from multivariable calculus. In K. Cline & H. Zullo (Eds.), *Teaching Mathematics with Classroom Voting: With and Without Clickers*, MAA Notes (Vol. 79, pp. 127-130). Washington, DC: Mathematical Association of America.

Holly Zullo is an associate professor of mathematics at Carroll College in Helena, MT, and has used classroom voting in calculus, linear algebra, and differential equations for several years. She is co-PI on the NSF-funded grant, Math Vote, a 3-year grant to study the pedagogy of classroom voting.

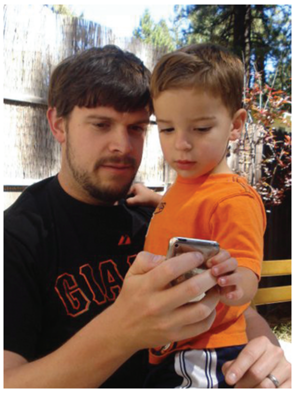
Jean McGivney-Burelle is an associate professor of mathematics at the University of Hartford. She received her MS from Northeastern University and her PhD from the University of Connecticut. Her research interests involve investigating how to use technology to improve the teaching and learning of mathematics.

Kelly Cline is an associate professor at Carroll College in Helena, MT, where he has used classroom voting in calculus, linear algebra, and differential equations since 2005. He is currently PI on Math Vote, an NSF funded three-year project to study the pedagogy of classroom voting.

Ann Stewart is an assistant professor of mathematics at Hood College in Frederick, MD. She uses classroom voting in many of her courses, and she is also interested in connections between mathematics and music. She serves as senior personnel on the NSF-funded project *Math Vote*.

Chris Storm has been an assistant professor in the Department of Mathematics and Computer Science at Adelphi University since fall 2007. He did his graduate work at Dartmouth College. He is an Exxon-Mobil Project NExT Sun Dot (2007) Fellow. In his free time, he enjoys playing bridge and hiking.

David A. Huckaby is an associate professor of mathematics at Angelo State University. He received his BS degree from the University of Texas at Austin and his MA and PhD in applied mathematics from UCLA.



Tim Melvin received his BS in mathematics from Carroll College, his MA in mathematics from Sacramento State University and is currently working on his PhD in mathematics at Washington State University. His professional interests include linear algebra, commutative algebra, and mathematical foundations.