CAN GOOD QUESTIONS AND PEER DISCUSSION IMPROVE CALCULUS INSTRUCTION?

Robyn L. Miller[1], Everilis Santana-Vega[2], and Maria S. Terrell [3]

ADDRESS: Department of Mathematics, Cornell University, Ithaca New York 14853 USA. (1) rlm35@cornell.edu, (2) ESantana@nhcc.edu, (3) mst1@cornell.edu.

ABSTRACT: Preliminary report of the results of a project[1] to introduce peer instruction into a multi-section first semester calculus course taught largely by novice instructors. This paper summarizes the instructional approaches instructors chose to use, and the subsequent results of student performance on common exams throughout the course of the term.

KEYWORDS: Calculus, good questions, in class polling, peer instruction, ConcepTests, student achievement, teaching assistant development,interactive engagement, under represented minorities.

INTRODUCTION

Can you raise the visibility of key calculus concepts, promote a more active learning environment, support young instructors in their professional development in their early formative teaching experiences, and improve student learning? We think the answer is yes, if you ask students Good Questions and encourage them to refine their thinking with their peers. What makes a question good?

[1] Support for the Good Questions project was provided by the National Science Foundation's Course, Curriculum, and Laboratory Improvement Program under grant DUE-0231154. Opinions expressed are those of the authors and not necessarily those of the Foundation.

Imagine a classroom where the instructor pauses every fifteen minutes or so to ask a highly conceptual multiple choice or True-False question. For example

True or False: You were once *exactly* π feet tall.

Students think about the question independently and register their vote. As the instructor uses that feedback to start to assess the state of the class's understanding, students are encouraged to discuss their answers with someone sitting near by, preferably a student who is thinking about the problem differently. As the room erupts with inquires of "what did you think?" and "why did you think that?", and with replies of "well I'm not sure, but I think...", the instructor listens in on conversations. Students share their reasoning, argue its validity, and work together as they think more deeply about what the question means, and how the mathematical ideas, definitions, and or theorems from their text or lecture might apply. The instructor calls for a revote, shares the result with the class, and asks for a student who changed their mind to share with the class what caused them to change their mind. In the space of four to five minutes the instructor has provoked students to think, surveyed students' votes, listened to students' conversations, and obtained valuable insight into students' thinking about key concepts. During that same time students have wondered, conjectured, reasoned, argued, justified, probed mathematical concepts, refined their mathematical reasoning and understanding, and shared that new understanding with the class. This is what we call a "Good Questions" approach with peer discussion.

Good Questions for teaching mathematics is adapted from Eric Mazur's [3] *ConcepTests and Peer Instruction* for teaching physics. There is a significant body of research and results in physics education that show that interactive engagement in undergraduate physics instruction leads to higher student achievement. (See Hake [2] for a recent overview and references.) Crouch and Mazur's [1] work shows that well constructed multiple choice questions, used in conjunction with pre-class warmup reading assignments, can engage students in productive peer discussions that enhance student learning. In adapting Mazur's work to teaching calculus we learned (Santana-Vega [6]) that a well constructed multiple choice question and the ensuing peer discussion it can provoke is a powerful teaching tool.

Our project was not the first to adapt Mazur's approach to interactive engagement to the calculus classroom. But our questions are somewhat different than those developed by Scott Pilzer [4] who developed ConcepTests for calculus early on and then expanded upon that work with others from the Calculus Consortium. (See Pilzer *et al.* [5].) A full collection of our questions in downloadable forms can be found at our project website `http://www.math.cornell.edu/~GoodQuestions`

and at the Canadian in-Class Questions Database

`http://cinqdb.physics.utoronto.ca`

We tested Good Questions (GQ) Fall 2003 in traditional first semester calculus at Cornell University. There were 330 students enrolled in 17 small sections of 15 to 26 students. These small sections were taught by 14 different instructors who had the freedom to choose their teaching method, including whether or not to use GQ. Class met 4 days a week for 50 minutes with the instructor. We used electronic polling devices, student surveys, instructor surveys, and exam performance throughout the term to collect data about which questions were used, how they were used, and what effect using questions had on student performance.

But what do we mean by a good question? Aren't all questions that test student knowledge and mastery of the material equally "good"? One might think for example, that a good question should be unambiguously clear, have a unique correct solution, and be framed with perfect mathematical precision. We found that the best questions, as measured by how frequently instructors chose to use them and on how successful questions were in stimulating class discussion, were those that had more than one interpretation, or had more than one or perhaps no solution. Questions like that were somehow more discussable.

For example, in the question about being π feet tall, there is usually a reliable but small minority, about 20-25%, who shock their peers by voting "False". False votes appear even more shocking if the question is asked, "true or false you were once exactly 3 feet tall". When encouraged to explain their reasoning, those who vote false usually convince their classmates that they have a valid concern. They explain that they believe that growth happens in a discrete way, a molecule at a time. Therefore, they may have skipped over the π foot mark. In the end the discussion centers on whether the hypothesis of the theorem is satisfied. This question not only gives a teacher the opportunity to stimulate a good discussion of the intermediate value theorem in an intuitive setting, but it provides a context to discuss the simplifying assumptions we often make in assuming continuity, and the challenges of building a mathematical model.

A good question can help teachers and students probe and reveal common misunderstandings and misconceptions. For example, we found some students vote "False" even though they see growth as a continuous function of time. These students argue that they can never be exactly π feet tall because π is irrational. Their justification runs something like this, "because

π is irrational, it cannot be precisely measured, therefore you can never be *exactly* π feet tall". This kind of reasoning usually brings responses from other students that the intermediate value theorem is an existence theorem, and once growth is assumed to be continuous, there had to be a time that they were π feet tall even if they don't know π's exact location on a ruler. A good question invites students to discuss the mathematics they are learning, what it means to them, and occasionally it provokes deeper discussion of fundamental concepts and ideas.

Good questions draw on students prior knowledge and ways of thinking and connect it to the mathematics they are learning. For example, students naturally approximate e^2 by squaring numbers close to e. The next question invites students to link this intuitive and practical notion of continuity to the formal definition of continuity. It also helps students understand the nature of real numbers and why continuous functions are so highly prized.

> You decide to estimate e^2 by squaring longer decimal approximations of $e = 2.71828\ldots$.
>
> (a) This is a good idea because e is a rational number.
> (b) This is a good idea because $y = x^2$ is a continuous function.
> (c) This is a bad idea because e is irrational.
> (d) This is a good idea because $y = e^x$ is a continuous function.

We found that on the first polling, generally 60-70% of students chose (d),15-20% chose (b) which is correct, 10-15% chose (c), and 5% chose (a). In the course of peer discussion as students try to articulate why they think (d) is the correct response, they find themselves looking for connections between the formal definition of continuity and sketches of $y = e^x$. What they find is that the process that the problem describes involves sneaking up on e, not 2. After approximately 3-4 minutes of peer discussion, the second polling usually shows that better than 80% of the students choose (b). In general this large shift in correct responses is not due to students with the correct response convincing others; it is students with incorrect responses clarifying their thinking and making the discovery themselves. Connecting their intuition with the formal mathematical language gives students a precise way of discussing a phenomenon they have experienced intuitively, that you can get a number as close as you like to e^2 by squaring a number that is close enough to e. A Good Question provokes students to use the language of mathematics in natural ways to describe processes and procedures that are already natural and intuitive to them. In this way the concepts we are defining with mathematical language and symbols take on a refining and clarifying role.

A good collection of questions needs to include a wide variety of types of questions for different purposes. The next two questions are more traditional and have been successful in revealing student misunderstandings about what a "tangent line" is in the context of calculus.

At the point $(0, 0)$ the graph of $y = |x|$

(a) has exactly one tangent line, $y = 0$.
(b) has infinitely many tangent lines.
(c) has two tangent lines, $y = x$ and $y = -x$
(d) has no tangent line.

On the first vote when used shortly after introducing the notion of tangent line in calculus, class responses frequently are: a) 20%, b) 45% , c) 10%, d) 25%. By listening to students' discussions we have found that even when students have learned that the derivative of this function does not exist at x=0, students have a strongly held belief that a tangent line is a line that touches at one point. As a result, a) and b) are very popular, c) captures some guessers and a very small group of students who essentially think of tangent lines as support lines, or a kind of one-sided local linearity. The correct answer (d) does gain in favor as students discuss their ideas. The most common arguments go as follows: "A tangent line at $(0, 0)$ needs to have a slope, that is the derivative at $(0, 0)$. The derivative does not exist at a corner point (a fact students seem to know). As a result, there cannot be a tangent line at $(0, 0)$." This line of reasoning is much more procedural and less intuitive than an argument that relies on the idea of "local linearity". In our tests of this question we rarely hear students refer to the tangent line as the line the graph resembles if you zoom in close enough. The next Good Question helps to raise the visibility of "local linearity" and to confront the danger in the "touching at one point" interpretation.

At the point $(0, 0)$ the graph of the function $f(x) = x$

(a) has exactly one tangent line, $y = 0$.
(b) has infinitely many tangent lines.
(c) has exactly one tangent line $y = x$
(d) has no tangent line.

On the first vote when used immediately after the question above, class responses frequently are: a) 5%, b) 10%, c) 30%, d) 55%. The ensuing peer discussion usually results in a large shift, if not unanimous agreement to the correct response, c). Here again students reason in a very formal way, they compute the derivative, $f'(x) = 1$. They use the slope and the point $(0, 0)$ to "get the equation of the tangent line", $y = x$. These Good Questions help instructors see how difficult it is for many students to adopt the local linearity notion of tangent line. Such an interpretation of the tangent line would have allowed them to see immediately that a linear function is identically equal to its tangent line at every point. A good question is a helpful tool for giving an instructor insight into how students reason about the mathematics they are learning.

Our collection includes a number of Good Questions that generate very deep discussions about what the derivative is by asking students to interpret the derivative in a physical context that is natural, but non-numerical and non-graphical. For example:

> As you cut slices off a loaf of bread the volume changes as you change of the length of the loaf. The derivative of the volume with respect to the length is
>
> (a) the surface area of the part of the loaf that's left.
> (b) the area of the cut face of the loaf.
> (c) the volume of the last slice divided by the thickness of the last slice.
> (d) the volume of the last slice you cut.

Students realize that the volume is changing as the length of the loaf changes, but they are unaccustomed to thinking about rate of change in physical terms that do not depend on time. Many students guess a) and b), but they are not at all sure. They know there is a relationship between the derivative of volume and area for boxes and spheres, but a loaf of bread? A few think that c) looks close (and they are right). During the discussion, some students begin to see that the change in the volume is the volume of the last slice, and that the change in the length is the thickness of that slice. What is very difficult for them to notice on their own is that that difference quotient is trapped between the largest and smallest cross sectional area of the slice, and that as the thickness of the last slice approaches 0, the difference quotient is trapped between two numbers that are approaching the same number, the area of the cut face of the loaf. This problem remains a mystery to students for quite some time. However, it is a very physical way for students to see that two quantities that are approaching zero, the change in the volume (i.e., the volume of the slice) and the change in the length (i.e., the thickness of the slice), have a ratio that is approaching a number they can see–the cross sectional area of the loaf at the place where you removed the slice.

The bread question is part of a sequence of questions we have about cutting things up and asking what the derivative is. For example, when you cut a slice of pizza, the size of your slice changes as a function of the angle you make when you make the second cut. What is the derivative of that function? If you slice your pizza by chopping it with a cleaver in parallel slices along its length (like the loaf of bread) what's the derivative of that function? If you decide to eat your pizza by cutting off concentric rings of pizza, i.e by changing its radius, what's the derivative of that function? If your pizza is circular, the answers are fairly easy. If your pizza is amoeba shaped, the responses require much deeper thought.

SOME CHARACTERISTICS OF A GOOD QUESTION

- Stimulates students' interest and curiosity in mathematics.
- Helps students monitor their understanding.
- Offers students frequent opportunities to make conjectures and argue about their validity.
- Draws on students' prior knowledge, understanding, and/or misunderstanding.
- Provides instructors a tool for frequent formative assessments of what their students are learning.
- Supports instructors' efforts to foster an active learning environment.

SUMMARY OF PRELIMINARY FINDINGS

What did we learn about how instructors used GQ? What kinds of questions did instructors choose to use?

Good Questions were tagged with three labels (Quick Check, Probing, and Deep) that reflected that questions were designed to assess and to engage students in progressively deeper levels of mathematical thinking and reasoning. Most instructors most frequently used Quick Checks in their teaching, but two instructors primarily used Probing and Deep questions.

How frequently did instructors use GQ?

We surveyed instructors and students three times in the term to determine how frequently and in what ways Good Questions were being used in class. Generally the surveys were consistent, but where instructors and students gave different reports on the use of peer discussion, we relied on the students' assessment.

Instructors' patterns of use of GQ fell into four groups:

- *Deep*– Deep and Probing questions at least 1, but usually 2 or 3 days each week, and used 2 or 3 questions per class, with peer discussion.
- *Heavy plus Peer*– Primarily Quick Check questions 3-4 days per week, and used 2 or 3 questions per class with peer discussion;
- *Heavy no Peer*– Primarily Quick Checks 3 or 4 days each week, and used 2 or 3 questions per class, with little or no peer discussion.
- *Light to Nil* Used GQ sporadically if ever, and used little or no peer discussion.

What kind of data on student performance did we gather?

- *Regular course-wide common exams*– 3 regular 100 point exams (called preliminary or Prelims) and one 150 point comprehensive final exam. Students papers were graded consistently across all classes. Each preliminary exam had questions identified as conceptual. We recorded conceptual subscores in addition to total scores.
- *SAT and demographic data*– were provided to us by university administration.

All Students

Variable	No Peer Means	N	Peer Means	N	Null: Means Equal? ($\alpha = 0.05$)	Level of Signif.
Prelim1	82 (B)	136	86 (B+)	97	Reject	0.0029596
Conceptual Subscore	20	135	21	97	Reject	0.04052
Prelim2	76 (B-)	153	80 (B)	96	Reject	0.0040231
Prelim2 Con.	18	153	20	96	Reject	0.0011933
Prelim3	72 (C+)	152	77 (B-)	96	Reject	0.0040928
Prelim3 Con.	32	148	35	96	Reject	0.013646
Final Exam	117 (B-)	153	126 (B)	96	Reject	0.00018824

Table 1. 1-sided T-Tests for Difference in Means Between "Peer" and "No Peer".

Women

Variable	No Peer Means	N	Peer Means	N	Null: Means Equal? ($\alpha = 0.05$)	Level of Signif.
Prelim1	82 (B)	76	86 (B+)	57	Reject	0.015822
Prelim1 Con.	19	76	21	57	Accept	0.056766
Prelim2	76 (B-)	86	79 (B)	56	Reject	0.025198
Prelim2 Con.	18	86	20	56	Reject	0.028832
Prelim3	73 (B-)	85	76 (B-)	57	Accept	0.10499
Prelim3 Con.	33	83	34	57	Accept	0.23062
Final Exam	115 (B-)	86	125 (B)	56	Reject	0.0023054

Men

Variable	No Peer Means	N	Peer Means	N	Null: Means Equal? ($\alpha = 0.05$)	Level of Signif.
Prelim1	83 (B)	60	87 (B+)	40	Reject	0.041932
Prelim1 Con.	21	59	22	40	Accept	0.19529
Prelim2	76 (B-)	66	80 (B)	40	Reject	0.032915
Prelim2 Con.	18	66	21	40	Reject	0.005643
Prelim3	71 (C+)	66	78 (B-)	39	Reject	0.0058032
Prelim3 Con.	31	65	35	39	Reject	0.0092971
Final Exam	118 (B-)	66	126 (B)	40	Reject	0.015522

Under Represented Minorities

Variable	No Peer Means	N	Peer Means	N	Null: Means Equal? ($\alpha = 0.05$)	Level of Signif.
Prelim1	78 (B-)	18	81 (B)	22	Accept	0.21859
Prelim1 Con.	17	18	19	22	Accept	0.28009
Prelim2	69 (C)	24	72 (C+)	22	Accept	0.29334
Prelim2 Con.	17	24	17	22	Accept	0.4776
Prelim3	67 (C)	24	71 (C+)	21	Accept	0.19238
Prelim3 Con.	30	24	31	21	Accept	0.32179
Final Exam	106 (C)	24	117 (B-)	21	Reject	0.037757

Table 1. (continued) 1-sided T-Tests for Difference in Means Between "Peer" and "No Peer".

CONCLUDING REMARKS

Data suggest that the benefit of using Good Questions comes from the peer discussion, and that discussions about deeper questions may benefit students more. The greatest effect appeared on the comprehensive final exam for all students and all groups. Under represented minorities showed the largest numerical difference means on the final exam, 11 points. The data also suggest that for some students the benefit from GQ and peer discussion may not result in higher performance on the most conceptual questions, but that a better understanding of concepts enables them to improve their performance in the traditional parts of the course.

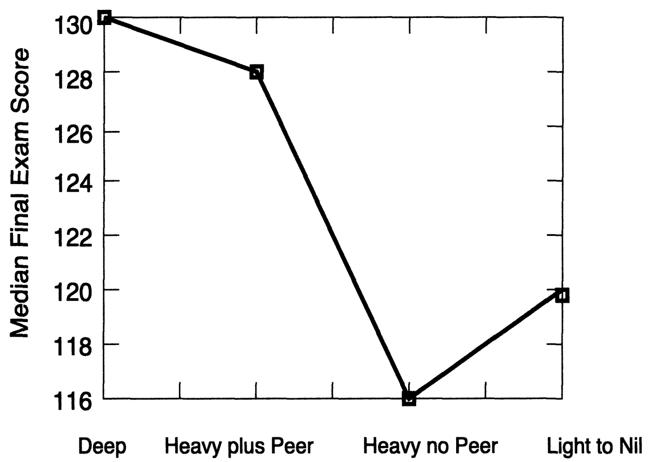

Figure 1. Final exam medians for different patterns of GQ use.

REFERENCES

1. Crouch, C. H. and E. Mazur. 2001. Peer Instruction: Ten years of experience and results. *Am. J. Phys.* 69: 970-977

2. Hake, R. 2005. The Physics Education Reform Effort: A Possible Model for Higher Education? *The National Teaching and Leaning Forum.* . 15 (1) `http://www.ntlf.com/html/ti/toc.htm`

3. Mazur, E. 1997. *Peer Instruction: A User's Manual.* Englewood Cliffs NJ: Prentice Hall.

4. Pilzer, S. 2001. Peer Instruction in Physics and Mathematics. *PRIMUS*. 11(2): 185-192.

5. Pilzer, S., M. Robinson, D. Lomen, D. Flath, D. Hughes Hallet, B. Lahme, J. Morris, W. McCallum, J. Thrash. 2003. *ConcepTests to*

Accompany Calculus, Third Edition. Hoboken NJ: John Wiley & Son.

6. Santana-Vega, E. 2004. The Impact of the *Good Questions Project* on Students' Understanding of Calculus Concepts. MS Thesis Cornell University, Ithaca NY.

BIOGRAPHICAL SKETCH

Robyn Miller is a graduate student in Mathematics at Cornell University.

Everilis Santana-Vega teaches at North Hennepin Community College in Brooklyn Park MN. She earned a Masters Degree in Mathematics at Cornell University under the advice of Dr. David W. Henderson. She is interested in undergraduate mathematics education.

Maria Terrell is a Senior Lecturer in Mathematics at Cornell University, where she is currently Director of Teaching Assistant Programs. She was a student of Edwin Floyd at the University of Virginia, where she earned a PhD in algebraic topology. Her current research interests include the geometry of rigid structures, undergraduate mathematics education, and professional development for teaching assistants in mathematics.