ConcepTest Response Times in Peer Instruction Classrooms

Kelly Miller,[1] Nathaniel Lasry,[2] Brian Lukoff,[1] Julie Schell,[1] and Eric Mazur[1]

[1]*Department of Physics and Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA**

[2]*Department of Physics, John Abbott College, Montreal, Quebec H9X 3L9, Canada*

(Dated: September 22, 2014)

Classroom Response Systems (CRSs) are widely used in interactive teaching environments as a way to engage students by asking them questions. Previous research on the time taken by students to respond to conceptual questions has yielded insights on how students think and change conceptions. We measure the amount of time students take to respond to in-class, conceptual questions (ConcepTests or CTs) in two introductory physics courses taught using Peer Instruction and use Item Response Theory to determine the difficulty of the CTs. We examine response time differences between correct and incorrect answers both before and after the peer discussion for CTs of varying difficulty. We also determine the relationship between response time and students performance on a standardized test of incoming physics knowledge, pre-course self-efficacy, and gender. Our data reveal three results of interest: First, response time for correct answers is significantly faster than for incorrect answers, both before and after peer discussion, especially for easy CTs. Second, students with greater incoming physics knowledge and higher self-efficacy respond faster in both rounds. Third, there is no gender difference in response rate after controlling for incoming physics knowledge scores, although males register significantly more attempts before committing to a final answer than do female students. These results provide insight into effective CT pacing during Peer Instruction. In particular, in order to maintain a pace that keeps everyone engaged students should not be given too much time to respond. Once around 80% of the answers are in, the ratio of correct to incorrect responses rapidly approach levels indicating random guessing and instructors should close the poll.

INTRODUCTION

When used effectively, Classroom Response Systems (CRSs) can facilitate student engagement in science classrooms in ways that would not be possible otherwise[1, 2]. From low-tech versions, such as flashcards or white boards, to higher-tech, web-based response systems, such as Poll Everywhere or Learning Catalytics, the popularity of CRSs has also opened up an active area for educational researchers to pursue questions about their use and student learning outcomes. Peer Instruction is one popular research-based instructional strategy that leverages the power of CRSs to promote student learning [3]. In Peer Instruction, students use CRSs to respond to ConcepTests (CTs) both before and after discussing their answers with their peers for 2 to 5 minutes [4]. ConcepTests are short conceptual questions that focus on a single topic [3]. During Peer Instruction, students first respond individually to a CT and then respond a second time to the same CT after discussing their responses with a peer. Instructors can use CT response data to inform teaching decisions in real time. Researchers suggest that Peer instruction is a superior teaching strategy in promoting student conceptual understanding and problem solving skills compared to the traditional lecture [3–8]. The purpose of this research is to gain insight into student thinking in Peer Instruction environments using one specific unit of analysis related to CTs: response time.

Despite the fact that Peer Instruction is a research-based pedagogy shown to be highly effective in promoting student learning, there is room for improvement particularly in implementation. Response times can provide insight into student thinking during Peer Instruction. Previous research has shown that when students answer conceptual questions with misconception-like responses, they tend to respond more quickly than those answering correctly [9] With this insight, instructors can improve the implementation of Peer Instruction. For example, an examination of response time could inform CT pacing and help determine the optimal amount of time instructors should keep questions open for student response. Providing instructors with guidelines for appropriate CT pacing allows them to implement Peer Instruction more efficiently by optimizing class time.

We pose two research questions. First, what are the response time differences between correct and incorrect answers, before and after peer discussion for CTs of varying difficulty? Second, what is the relationship between response time and students performance on a standardized test of incoming physics knowledge, pre-course self-efficacy, and gender?

Physics education researchers have previously conducted studies of response times to conceptual questions to gain insight into student learning. Using standardized conceptual questions delivered as a pre-post course conceptual survey, Lasry et al. (2013) showed that response times for incorrect answers are longer than for correct responses [10]. Our study differs from Lasry et al. (2013) in that we examine in-class formative conceptual questions versus pre-post course surveys. Richardson et al. (2013) examined the relationship between gender and CT response time in a Peer Instruction classroom and found no statistically significant difference in response time between males and females [11]. However, Richardson et al. (2013) also found that males were more likely than females to change their answer within a single round before committing to a final response [11]. In this study of CT response times, we extend this work by combining response times with student learning data to gain further insight into student thinking in Peer Instruction environments.

METHODS

We collected student responses and response times for ConcepTests from one semester in two introductory electricity and magnetism classes at Queens University (Kingston, Ontario; N = 48) and Harvard University (Cambridge, Massachusetts; N = 93). Both classes used Peer Instruction with between five and 15 CTs per class. Over the course of the semester, 101 CTs were given at Queens and 74 at Harvard. In both classes, students did not receive credit for getting the answer right. Credit was only given for responding to the questions. Students answered each CT in two rounds of questioning by entering their responses via a classroom response system (CRS) (iClicker at Queens University and Learning Catalytics at Harvard). The CRS recorded a timestamp for each students final response. Response times were computed as the difference between the time the question was delivered to the students (via the CRS) and the time of each students final response. Regardless of the number of responses a student entered for a single question, only the time to the final response was recorded. For each question, we analyzed the time taken to respond before and after the discussion. The response time before the discussion is the amount of time it takes students to respond individually to the question the first time they see it. After students discuss answers with each other, the instructor presents the question again and opens the polling for a second time. The response time after the discussion is the amount of time that students take to respond after the polling is opened the second time. We did not constrain how long each question would stay open because some questions take longer to parse than others. We also chose to give each instructor the control of how much time each question stayed open. Students often continue their discussion past the time when the poll is reopened for the second round. Consequently, some questions stayed open much longer than others. In some rare cases, students were given more than five minutes to respond to a single question, especially in the second round. To control for the variability in response times across questions, we focus our analysis on the difference in the time taken to respond correctly and incorrectly. The iClicker system used at Queens also records the number of times a student responds to a CT before registering their final answer and, for this subset of the data, we included this in our analysis.

The Conceptual Survey of Electricity and Magnetism (CSEM) [12] and a self-efficacy survey were both administered twice in each class, once as pre-tests at the beginning of the semester and again as post-tests at the end of the semester. Self-efficacy is a person's belief that s/he can be successful when performing a situation-specific task [13], and students' self-efficacy has been shown to be a strong predictor for perseverance and success in science [14]. The self-efficacy survey used in this study was developed at Harvard, and based on the Sources of Self-efficacy in Science Courses (SOSESC) survey [14]. This survey asked students to rank, on a five-point scale, the extent to which they believe they will be successful in a number of physics-related tasks (e.g., how successful will they be solving difficult physics problems or communicating physics successfully to a peer).

A two-parameter Item Response Model [15] was applied to the first round of responses across all items to determine the difficulty of each item (b-parameter). Items were grouped into two categories according to difficulty. Items for which $b < 0$ were classified as easy and items with $b \geq 0$ were classified as hard. In the Queens dataset, 59 out of the 101 CTs were classified as easy and the remaining 42 as hard. In the Harvard dataset, 38 of the 74 CTs were classified easy and the remaining 36 as hard.

RESULTS

Table 1 shows the average time taken for correct and incorrect answers on ConcepTests both before $\langle t_{before} \rangle$ and after $\langle t_{after} \rangle$ peer discussion at Queen's and Harvard University (QU and HU respectively). Before peer instruction, the average response time is 20-30% shorter for correct answers than for incorrect answers ($p < 0.001$). After peer discussion, the average response time is about 10% shorter for correct answers than for incorrect answers ($p < 0.01$).

We also analyzed the average response time difference between incorrect and correct responses $\langle t_{incorr} \rangle$ - $\langle t_{corr} \rangle$ for each individual question. Of the 175 questions posed in this study: 27 had a statistically significant difference (at $p < 0.05$) where $\langle t_{incorr} \rangle > \langle t_{corr} \rangle$, three questions had a statistically significant difference (at $p < 0.05$) where $\langle t_{incorr} \rangle < \langle t_{corr} \rangle$ and 145 had no significant difference between $\langle t_{incorr} \rangle$ and $\langle t_{corr} \rangle$. Therefore, for the majority of the questions with a significant time difference, incorrect responses are not misconception-like in nature because students take longer to provide an incorrect answer than the correct one.

Figure 1 shows histograms for responses as a function

TABLE I. Average time taken for correct and incorrect answers on ConcepTests, before and after Peer Instruction.

	$\langle t_{before} \rangle\,(s)$		$\langle t_{after} \rangle\,(s)$	
	HU	QU	HU	QU
Correct	70.9	26.9	95.3	64.3
Incorrect	75.8	35.5	102.3	69.9
Difference	4.9*	8.6**	7.0*	5.6*

**p <0.0001 * p <0.005

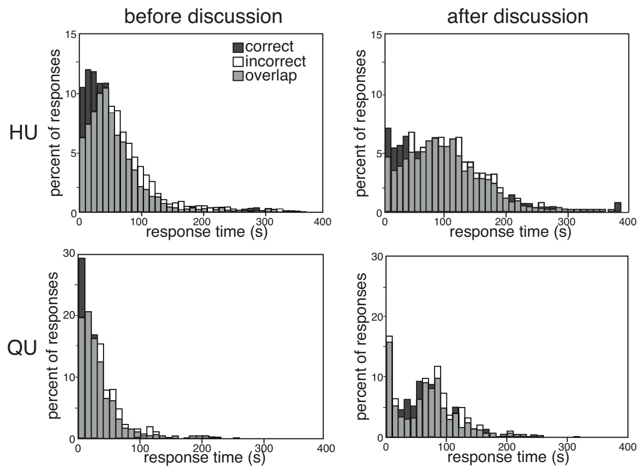

FIG. 1. Distributions of ConcepTest response times before and after peer discussion.

of ConcepTest response time before and after peer discussion. Correct responses (dark grey) and incorrect responses (white) are expressed as a percentage of all responses over all the ConcepTest questions posed during the semester. The light grey area represents the overlap of the white and dark grey bars. The top two histograms represent data from Harvard (HU) and the bottom two histograms represent data from Queen's (QU). Figure 1 illustrates the shorter time scale with which correct responses are given compared to incorrect responses. Given that there was no set amount of time that the polling stayed open, we find some questions were left open for more than 5 minutes.

Figure 2 shows the ratios of correct to incorrect responses plotted as a function of the percent of students in the class who have entered their response. Ratios of correct to incorrect responses are shown for responses before the discussion (pre) and after the discussion (post) for questions from both the Harvard classroom (HU) and the Queens classroom (QU). Error bars represent the standard error of the mean across each of the four ratios. Figure 2 shows that as more students respond, the proportion of correct answers decreases compared to the proportion of incorrect answers, both before and after the peer discussion. Once about 80% of the students in the class have responded, the ratio of correct to incorrect answers rapidly approaches levels indicating random guessing by the students.

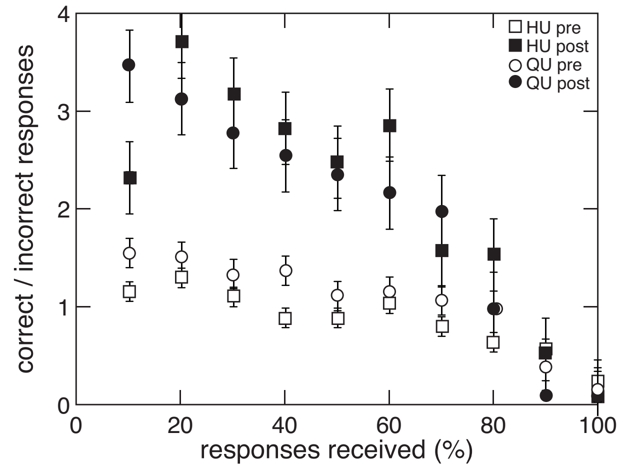

FIG. 2. Ratio of correct to incorrect responses as a function of the percentage of responses received.

TABLE II. Average time taken for correct and incorrect answers on ConcepTests classified as easy and hard, before and after peer discussion

	$\langle t_{before} \rangle$ (s)				$\langle t_{after} \rangle$ (s)			
	HU		QU		HU		QU	
	Easy	Hard	Easy	Hard	Easy	Hard	Easy	Hard
Correct	52.9	65.0	24.9	35.3	80.5	109.9	52.2	87.9
Incorrect	62.2	69.9	29.1	36.1	101.8	113.9	60.4	89.3
Difference	9.3**	4.9	4.2**	0.8	21.3**	4.0	8.2**	1.4

**p <0.0001 * p <0.005

Table 2 displays the average response times for correct and incorrect answers for questions determined to be easy and hard. The difference in response time was determined by subtracting the average response time for correct answers from the average response time for incorrect answers. The difference is positive for easy and hard questions both before and after the peer discussion indicating that, regardless of the difficulty of the question or whether the response is provided before or after the discussion, students take longer to respond with an incorrect answer than with a correct one. However, this response time difference is only statistically significant for easy questions (p <0.0001). For difficult questions, while the average response time is longer for incorrect answers than for correct answers, the difference is small and not statistically significant. Table 2 also shows that students take longer to answer hard questions than easy questions, regardless of whether they are answering correctly or incorrectly.

Figure 3 shows that students with more incoming physics knowledge and higher self-efficacy respond faster, both before and after the peer discussion. Figure 3 dis-

TABLE III. Regression models predicting student response times for correct and incorrect responses both before and after the discussion

	Before Discussion				After Discussion			
	Correct		Incorrect		Correct		Incorrect	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Gender	9.5*	2.38	3.71	1.16	8.5*	4.9	9.9**	5.12
Pre-course physics knowledge		-1.27***		-0.67*		-1.14***		-1.23**
Pre-course self-efficacy		0.58***		0.50***		0.39***		0.38***
R^2	0.15	0.14	0.15	0.14	0.14	0.11	0.16	0.15
RMSE	62.3**	65.4	55.5	56.7	59.7	59.6	57.4	57.9

*p <0.05 ** p <0.01 *** p <0.001

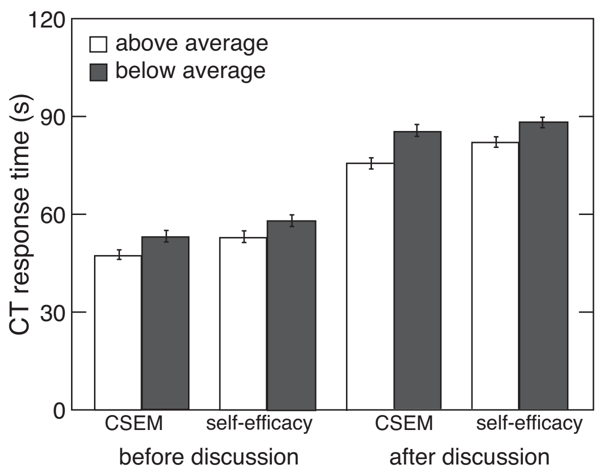

FIG. 3. Average response times before and after discussion for students with below and above average pre-course CSEM scores and below and above average pre-course self-efficacy scores.

plays average response times before and after the peer discussion for students in four different groups: (1) below average pre-course CSEM scores, (2) above average pre-course CSEM scores, (3) below average pre-course self-efficacy scores, and (4) above average pre-course self-efficacy scores. Students with higher incoming physics knowledge (above average pre-course CSEM scores) respond approximately 15% (p <0.001) faster before the peer discussion, and around 10% (p <0.001) faster after the peer discussion, than students with lower incoming physics knowledge (below average pre-course CSEM scores). Students with above average scores on the pre-course self-efficacy survey respond about 10% (p <0.001) faster both before and after the peer discussion, than students with below average scores.

Figure 4 shows male and female students response times for correct and incorrect answers. The first set of bars in each cluster (model 1) displays the response times when controlling for gender only while, the second set of bars (model 2) displays the response times

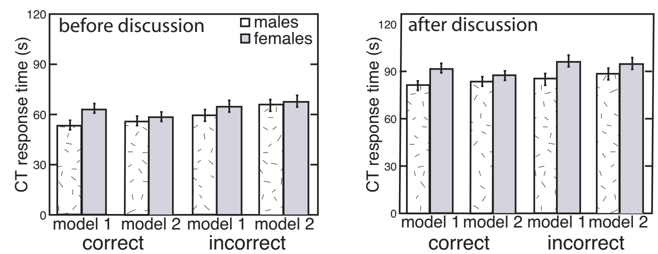

FIG. 4. Difference in response times both before and after discussion. We show both model 1, which controls only for gender and model 2, which also controls for pre-course CSEM and self-efficacy scores.

after adding pre-course CSEM and self-efficacy scores to model 1 as additional predictor variables. The left plot of Figure 4 indicates that males respond faster than females before discussion, regardless of whether they have the answer correct or incorrect, but that this difference disappears when students pre-course knowledge and self-efficacy are controlled for. The right plot of figure 4 shows that the same is true when considering male and female response times after the peer instruction. The regression parameters and significance metrics for both models are displayed in table 3.

Table 4 shows that a gender difference exists in the number of attempts students register before committing to a final answer. Compared to female students, male students respond significantly more times before deciding on a final response. The average number of attempts made by males exceeds that made by females both before $\langle n_{before} \rangle$ and after $\langle n_{after} \rangle$ the peer discussion. Before the peer discussion, male students change their initial response 40% more times than female students (p <0.001) before deciding on a final response. After the peer discussion, male students change their initial response 65% more times than female students (p <0.001) before deciding on a final response.

TABLE IV. Average number of attempts for males and females on ConcepTests, before and after the peer discussion

	$\langle n_{before} \rangle$	$\langle n_{after} \rangle$
Males	2.4	3.0
Females	1.7	1.8
Difference	1.3**	1.2**

***p <0.0001

DISCUSSION

A. Incorrect answers take more time

Multiple-choice ConcepTests are designed so that the incorrect choices are distractors, that is, incorrect ideas that are commonly held by students. These conceptions are often portrayed as common-sense beliefs that are stable and resistant to instruction [16]. Previous research has shown that when students answer conceptual questions with misconception-like responses, they tend to respond more quickly than those answering correctly [9]. In other words, strong distractors should yield quick responses. However, we find that students take longer to respond when they answer incorrectly, suggesting that these choices are not seen as strong distractors that yield automatic responses. Our data therefore suggests that when students select an incorrect answer, it is more likely because they do not know the answer, rather than because they are confident about the wrong answer.

We find that students take longer to respond to more difficult questions. Instructors frequently adjust their pacing based on the fraction of the class that has responded to a question, leaving the poll open longer for more difficult questions. Therefore, the student response rate for a question dictates the length of time the poll for that question is left open, but only up to a point. Instructors often close the poll before 100% of their class has responded though, according to our data, rarely when less than 77% has voted. On average, in the two classrooms used in this study, polls were closed once 91% of students had responded. Figure 2 shows that once more than 80% of students have responded, the ratio of correct to incorrect answers is so low that instructors should consider closing the poll. Once 90% of students have responded, the correct to incorrect ratio is, on average 25%, indicating that students are doing no better than random guessing.

We find that the difference in response time between correct and incorrect answers depends on the difficulty of the question. For easy questions, incorrect answers are not automatic; they take significantly longer than correct answers. Students answering easy questions incorrectly are spending time thinking. For harder questions, this does not appear to be the case. When the questions are difficult, the time taken to give an incorrect answer does not differ statistically from the time taken to give a correct answer. Apparently, the correct choice is not as obvious as for easy questions and students take as much time evaluating correct and incorrect options.

Response time for both correct and incorrect answers is significantly longer after peer discussion. Figure 1 shows that after discussing the question with their peers, students take longer to respond to the question than when they respond individually. There are two possible explanations for this finding. It is possible that response times are longer in the second round $\langle t_{after} \rangle$ because students take more time to think about the question. Alternatively, it could simply be due to the fact that some students may continue talking after the polling has reopened and that $\langle t_{after} \rangle$ includes some of the discussion time.

Our findings are based on CT response time data collected at two different institutions in two different classrooms with instructors who had very different pacing. A comparison of the top two histograms and the bottom two histograms in Figure 1 illustrates the different pace at which CTs were posed to students at Harvard (top) and to students at Queen's (bottom). On average, the polling in the Harvard classroom was kept open much longer than in the Queen's classroom, both before and after the discussion. Despite this difference in pacing, the same time scale difference emerges between correct and incorrect responses. Regardless of the institution and how long students are given to answer the questions, incorrect answers take more time than correct answers.

B. Response times vary according to students pre-course characteristics

Figure 3 shows that the students who respond faster are those who know more physics and have a stronger belief in their ability to be successful in a physics course at the beginning of the semester. These students spend less time considering alternative answers, that is, their answers are more automatic. Interestingly, these students answer more quickly, regardless of whether they are answering correctly or incorrectly. This finding suggests that response times are dependent on intrinsic student characteristics as well as cognitive processes.

C. CTs Responses and Gender

In contrast to earlier findings [11], we find that although males respond significantly faster than females, the difference disappears after controlling for self-efficacy and CSEM scores. The gender difference in response times appears to be at least partly attributable to a difference in pre-course knowledge and self-efficacy. When answering correctly, males respond 20% faster than females before (p<0.05) the discussion and 10% faster after the (p <0.05) discussion. When answering incorrectly, the difference is less pronounced. Before the discussion males answer incorrectly 6% faster than females (albeit this difference is not significant) and 12% faster after the discussion (p <0.05). However, when a linear regression model is used to control for pre-course knowledge and self-efficacy, gender ceases to be a significant predictor of response times and the gender difference disappears completely.

Our results also show that, compared to female students, male students respond significantly more times before deciding on a final response. This result is consistent with previous findings that males are more likely than females to answer with more attempts [17] and change their answer within a single round before committing to a final response [11].

CONCLUSION

For two different student populations, we find three interesting results from an analysis of response times to conceptual questions posed in class. The first is that incorrect answers take more time than correct answers, especially for easy questions. This suggests that incorrect responses result from students not knowing the answer, rather than from strongly held misconceptions. The second is that students with greater incoming physics knowledge and higher self-efficacy respond faster, indicating that response times are partly a function of student characteristics. Third, there is no gender difference in response rate when other student characteristics are controlled for. In light of these findings, we recommend that instructors terminate polls once 80% of the answers are in, because at that point an increasing fraction of students respond by random guessing.

ACKNOWLEDGMENTS

Several people contributed to the work described in this paper. EM & NL conceived of the basic idea for this work. KM, BL, and EM designed and carried out the study, and KM analyzed the results. JS and BL designed and implemented the self-efficacy survey used in this study. EM & NL supervised the research and the development of the manuscript. KM wrote the first draft of the manuscript; all authors subsequently took part in the revision process and approved the final copy of the manuscript. James Fraser and Anneke Timan provided the data from Queens University and were important advisors throughout the study.

* kmiller@harvard.edu

[1] D. Bruff, *Teaching with classroom response systems: Creating active learning environments.* (John Wiley & Sons, 2009).
[2] J. Caldwell, CBE-Life Sciences Education (2007), 10.1187/cbe.06.
[3] E. Mazur, *Peer Instruction A User's Manual* (Prentice-Hall, Upper Saddle River, NJ, 1997).
[4] C. H. Crouch and E. Mazur, American Journal of Physics **69**, 970 (2001).
[5] S. B. McKagan, K. K. Perkins, and C. E. Wieman, AIP Conference Proceedings **883**, 34 (2007).
[6] R. N. Cortright, H. L. Collins, S. E. Dicarlo, N. Ronald, and E. Stephen, Advances in physiology education **2**, 107 (2005).
[7] A. P. Fagen, The Physics Teacher **40**, 206 (2002).
[8] S. P. Rao and S. E. DiCarlo, Advances in physiology education **24**, 51 (2000).
[9] A. Heckler, T. Scaife, and E. Sayre, in *Proceedings of the 32 annual conference of the cognitive science society* (2010) pp. 139–144.
[10] N. Lasry, J. Watkins, E. Mazur, and A. Ibrahim, American Journal of Physics **81**, 703 (2013).
[11] C. Richardson and B. O'Shea, American Journal of Physics **81**, 231 (2013).
[12] D. P. Maloney, T. L. O'Kuma, C. J. Hieggelke, and A. Van Heuvelen, American Journal of Physics **69**, S12 (2001).
[13] A. Bandura, Psychological review **84**, 191 (1977).
[14] H. Fencl and K. Scheel, AIP Conference Proceedings , 173 (2004).
[15] D. Thissen and S. Lynne, Psychometrika **51**, 567 (1986).
[16] H. Halloun I.A. D., American Journal of Physics **53**, 1043 (1985).
[17] G. Kortemeyer, Physical Review Special Topics - Physics Education Research **5**, 010107 (2009).