The effect of grading incentive on student discourse in Peer Instruction

Mark C. James

Citation: American Journal of Physics **74**, 689 (2006); doi: 10.1119/1.2198887
View online: https://doi.org/10.1119/1.2198887
View Table of Contents: https://aapt.scitation.org/toc/ajp/74/8
Published by the American Association of Physics Teachers

The effect of grading incentive on student discourse in Peer Instruction

Mark C. James[a)]
Department of Physics and Astronomy/Center for Science Teaching and Learning, Northern Arizona University, Flagstaff, Arizona 86011-6010

(Received 31 October 2005; accepted 31 March 2006)

The use of Peer Instruction to enhance lectures in large enrollment introductory college science courses has become widespread. In this technique, learner responses to multiple choice questions posed by the instructor during lecture are recorded and displayed in real time by an electronic classroom response system (CRS). Peer Instruction takes place when learners are given time to discuss ideas with their neighbors before registering individual responses. Although much research has been done to study the impact of Peer Instruction on student learning and engagement, little is known about the dynamics of the peer discussions that occur just before students register responses to questions. The results of this study suggest that the grading incentive instructors adopt for incorrect question responses impacts the nature and quality of the peer discussions that take place. Two large enrollment college astronomy classes employing contrasting assessment strategies for CRS scores were observed. In the high stakes classroom where students received little credit for incorrect CRS responses, it was found that conversation partners with greater knowledge tended to dominate peer discussions and partners with less knowledge were more passive. In the low stakes classroom where students received full credit for incorrect responses, it was found that students engaged in a more even examination of ideas from both partners. Conversation partners in the low stakes classroom were also far more likely to register dissimilar responses, suggesting that question response statistics in low stakes classrooms more accurately reflect current student understanding and therefore act as a better diagnostic tool for instructors.
[DOI: 10.1119/1.2198887]

I. INTRODUCTION

Eric Mazur has demonstrated that student performance on traditional tests of student learning in large enrollment introductory physics courses can be significantly increased when lectures are enhanced with Peer Instruction.[1] In this cooperative learning technique, the instructor interrupts a traditional lecture to pose a pre-planned multiple choice type question to assess student's understanding of the current topic. An electronic classroom response system (CRS) is used to enable learners to key in question responses via individual wireless remote controls.[2] Peer Instruction takes place when learners are given time to discuss their ideas with neighbors before registering individual responses. Statistics for the class responses to each question are displayed at the end of an allotted time interval. By analyzing student comprehension on the basis of these statistics, the instructor can decide to either proceed to the next topic or provide further instruction on the current topic. Instructors may choose to include student CRS scores in the overall course grading scheme and may adopt a grading incentive within the CRS grade book which assigns 0 to 100% credit for incorrect responses to reward class participation.

Several studies have examined the use of Peer Instruction in large enrollment college physics classes, and it has been shown to increase student engagement and comprehension[1,3,4] and increase student pass rate.[5] Peer Instruction has also been shown to be effective in other disciplines.[6,7] The degree of student engagement depends partially on learner expectations regarding their roles as participants in a Peer Instruction lecture.[8]

Modes of assessment play an important role in determining the nature of student learning that is realized in college classrooms.[9] Dickinson and Flick[10] have depicted how a traditional assessment system that emphasizes scoring for correct answers in an introductory physics course could undermine pedagogical goals of instruction. Boud, Cohen, and Sampson[11] have argued that to realize the potential benefits of all forms of peer learning, assessment practices must be aligned with teaching practices.

The purpose of this study is to examine how assessment practices relating to CRS questions influence the nature of conversations and degree of participation that occurs during Peer Instruction.

II. METHODS

A. Sample and context

In the Spring of 2005 two astronomy professors at a mid-sized university were recruited to participate in this study. Each instructor was assigned a graduate assistant whose duties were to operate the CRS system and collect data for use in this study. Each instructor received a stipend to support the development of CRS questions to be integrated into each lecture. The instructors were free to adapt their use of the system to suit their teaching styles. The larger course, with an enrollment of 180, was a standard first-year introductory astronomy course for non-science majors. The smaller course, with an enrollment of 84, was designed for first-year non-science majors interested in a multidisciplinary approach to studying the issues surrounding space travel and the possibility of extraterrestrial life. Student demographics in the two classes were similar with 66% freshman in the larger class and 64% in the smaller class. Both classes were 56% male. There were 89% non-science majors in the larger class and 84% non-science majors in the smaller class.

B. Data sources and methods

Instructors posed three to five multiple choice CRS questions during each lecture throughout the semester. After each question was posed, students were provided time to discuss ideas with a neighbor before responding to questions using their CRS response pads. Data on student discourse that occurred in response to CRS questions were collected during three class periods in each course near the beginning, middle, and end of the semester. A sample of 28 participants from the larger class and 24 participants from the smaller class used audio recorders to tape their CRS conversations during these classes. Learners who were invited to participate in the study were selected at random from the class and were provided a chance to win an iPod as an incentive to take part. To characterize the type of discourse that occurred during the recorded conversations, each idea that a student articulated was classified according to one of ten categories adapted from Ref. 12. The ten categories were restating question elements, stating answer preference, providing justification for way of thinking, posing a question or idea for consideration, articulating a new question that emerged from conversation not directly related to correctness of original question, stating agreement with partner's idea, rephrasing partner's idea, stating disagreement with partner's idea, asking for clarification regarding a partner's idea, and expressing uncertainty. The following example illustrates how a typical student's talking turn was tabulated in three categories: "The answer must be either C or D [stating answer preference]. I know the ocean temp is rising because they are worried about melting ice [providing justification], but I can't remember whether it is from carbon dioxide [expressing uncertainty]."

All recorded student conversations were tabulated independently by two research assistants who had been trained to sort student ideas into the ten categories. A very high correlation of $r=0.922$ was observed between the independent tabulations.[13] Non-CRS exam scores from recorded participants were analyzed to establish a relation between discourse characteristics and student knowledge in each class. CRS responses from recorded partners were also analyzed to establish the degree of consensus in question responses that occurred after peer discussions.

III. FINDINGS AND DISCUSSION

The two instructors in this study had different philosophies of how CRS scores were to be included in student's overall course grades. The instructor of the larger enrollment course adopted a high stakes approach where the grading scheme motivated students to answer CRS questions correctly. A student's overall CRS score counted for 12.5% of the overall course grade in this approach and incorrect responses earned one-third the credit earned by a correct response. The instructor of the lower enrollment course adopted a low stakes approach where student CRS scores were determined solely by the degree of participation. A student's overall CRS score in this approach counted for 20% of the course grade and incorrect responses earned as much credit as correct responses.

An interesting statistical difference between the classes was observed in the degree that conversations were dominated by one member of a conversing pair. The number of ideas articulated by a particular student during a conversation divided by the total number of ideas put forth by the conversing pair was the student's fractional contribution to the conversation. The difference between the fractional contributions from the partners was termed conversation "bias." In a conversing pair where one partner's conversational contribution score was 80%, the other partner's contribution score would be 20%, and the pair's discourse bias score would be 60%. As illustrated in Fig. 1, it was found that the mean discourse bias among conversation partners was much higher in the high stakes classroom (mean $M=33.2\%$, standard deviation $\mathrm{s.d.}=30.1\%$) compared to the mean discourse bias score among partners in the low stakes classroom ($M=14.8\%, \mathrm{s.d.}=10.9\%$). The group difference in discourse bias was statistically significant[14] with $F(1,45)=7.612$ and $p=0.008$. The student discourse contribution was correlated to course grade ($r=0.595, p=0.000$) in the high stakes classroom, where no such correlation existed in the low stakes classroom ($r=0.126, p=0.404$).

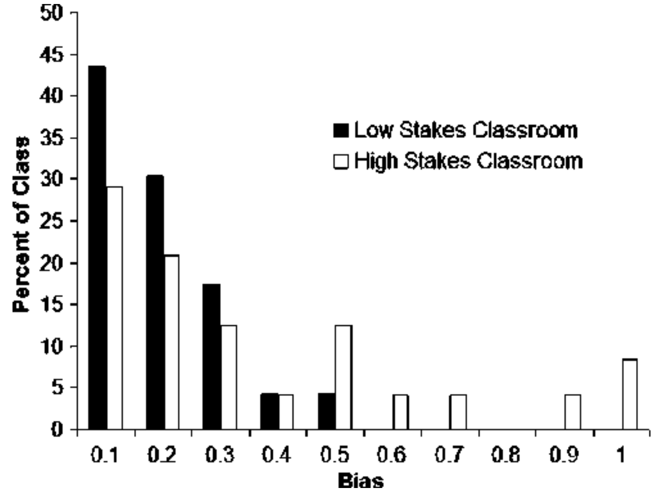

Fig. 1. Conversation bias indicating the balance between the ideas put forth by partners during Peer Instruction conversations.

In the high stakes classroom, students with more knowledge tended to be more dominant in CRS peer conversations, causing conversation bias to be large when the disparity between the partners' knowledge was large. Alternatively, the degree of conversation bias in the low stakes classroom was significantly lower and had no correlation to the degree of disparity of knowledge between partners. An analysis of conversation transcripts revealed that conversation partners with a large disparity in student knowledge in the high stakes classroom focused most of their discussion around the dominant student's answer preference. Such was not the case in the low stakes classroom where conversations were more balanced, including ideas put forth evenly from both partners.

The degree of consensus that was reflected in question responses among conversation partners was much higher in the high stakes classroom. The conversation partners in the high stakes classroom responded with a different answer to CRS questions in only 7.6% of responses, compared to 36.8% of contrasting question responses from partners in the low stakes classroom (see Fig. 2). The difference in partner consensus between classes was statistically significant with $F(1,166)=24.446$ and $p=0.000$. These results suggest that when there is a grading incentive that strongly favors correct responses to CRS questions, the question response statistics

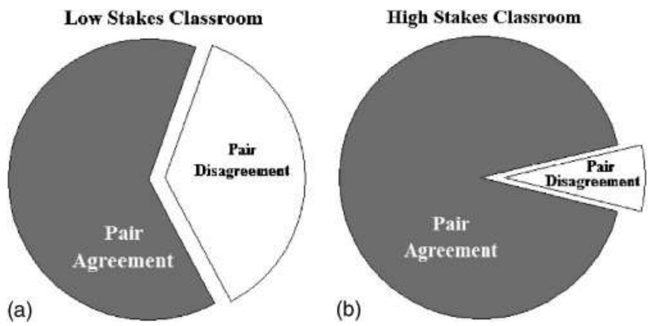

Fig. 2. Partner consensus indicating the degree of dissimilar responses registered by partners after Peer Instruction conversations.

displayed by the CRS system after each question may exaggerate the degree of understanding that actually exists and confound the ability of the instructor to make accurate pedagogical decisions based on student response feedback.

ACKNOWLEDGMENTS

Special thanks to NAU E-Learning Director Don Carter and Assessment Specialist Paula Garcia for their support of this study.

[a]Electronic mail: mark.james@nau.edu

[1]E. Mazur, *Peer Instruction: A User's Manual* (Prentice Hall, Upper Saddle River, NJ, 1991).

[2]Although the basic vote counting function described in this article is present in all CRS systems, vendors seek to distinguish their systems by including special options such as numeric entry, student confidence rating, LCD display on response pads, elaborate statistical packages for analyzing responses, seamless integration with POWERPOINT, and grade book synchronization with campus WEBCT and BLACKBOARD systems. Most systems, like the one used in this study distributed by eInstruction (⟨einstruction.com⟩), can assign 0 to 100% credit for incorrect CRS responses. Compact radio frequency based systems eliminate the troublesome installation issues found with the early line-of-sight infrared based systems and are now the industry standard. Most publishers offer discounted systems as an incentive for textbook adoption.

[3]D. E. Meltzer and K. Manivannan, "Transforming the lecture-hall environment: The fully interactive physics lecture," Am. J. Phys. **70**(6), 639–654 (2002).

[4]A. P. Fagen, C. H. Crouch, and E. Mazur, "Peer Instruction: Results from a range of classrooms," Phys. Teach. **40**, 206–209 (2002).

[5]J. Poulis, E. Robens, and M. Gilbert, "Physics lecturing with audience paced feedback," Am. J. Phys. **66**(5), 439–441 (1998).

[6]C. H. Crouch and E. Mazur, "Peer Instruction: Ten years of experience and results," Am. J. Phys. **69**(9), 970–977 (2001).

[7]S. P. Rao and S. E. DiCarlo, "Peer Instruction improves performance on quizzes," Adv. Physiology Educ. **24**, 51–55 (2000).

[8]L. A. Van Dijk, G. C. Van Der Berg, and H. Van Keulen, "Interactive lectures in engineering education," Eur. J. Eng. Educ. **26**(1), 15–28 (2001).

[9]L. C. Scharmann, M. C. James, and A. S. Smith, "Assessment in college science courses," in *Reform in Undergraduate Science Teaching for the 21st Century*, edited by D. W. Sunal and E. L. Wright (Information Age Publishing, Greenwich, 2004), Vol. 1, Chap. 8, pp. 137–152.

[10]V. L. Dickinson and L. B. Flick, "Beating the system: Course structure and student strategies in a traditional introductory undergraduate physics course for nonmajors," Sch. Sci. Math. **98**(5), 238–246 (1998).

[11]D. Boud, R. Cohen, and J. Sampson, "Peer learning and assessment," Assessment & Evaluation in Higher Educ. **24**(4), 413–426 (1999).

[12]S. Kaartinen and K. Kumpulainen, "Collaborative inquiry and the construction of explanations in the learning of science," Learn. Instr. **12**, 189–212 (2002).

[13]Pearson's r is a measure of the degree of linear relationship between two variables. A correlation of +1 would indicate a perfect positive linear relationship. Correlations above 0.9 are considered very high. The high correlation between tabulators in this study indicates that the criteria for tabulating conversations were well defined.

[14]The F test is a test for the statistical significance of an observed difference between the means of two samples. The numbers in parentheses indicate the number of groups minus one and the sample size minus the number of groups, respectively. The power of a test p is the probability that a statistical finding occurred by chance. A p value of 0.008 indicates less than a 1% probability that the observed difference in group means occurred by chance.