

# Astronomy Education Review

Volume 7, Feb 2008 - Dec 2008

Issue 1

## What Are They Talking About? Lessons Learned from a Study of Peer Instruction

by **Mark C. James**

Northern Arizona University

**Federica Barbieri**

Northern Arizona University

**Paula Garcia**

Northern Arizona University

Received: 01/11/08, Revised: 05/29/08, Posted: 07/07/08

The Astronomy Education Review, Issue 1, Volume 7:37-43, 2008

© 2008, Mark C. James. Copyright assigned to the Association of Universities for Research in Astronomy, Inc.

### Abstract

Electronic classroom response systems (CRSs) have been in use in large college lectures for over three decades. Such systems are designed to provide instructors and students with immediate statistical analyses of student electronic responses to multiple-choice questions posed to the class by the instructor. The technique known as peer instruction uses a CRS to tabulate student responses after students have had an opportunity to discuss ideas with seat partners. In this study, we investigate recorded peer CRS conversations collected in two introductory astronomy courses over two semesters. Findings suggest that when instructors adopt a high-stakes grading incentive that assigns little credit for incorrect CRS responses rather than providing forums for the spontaneous exploration of nascent ideas, conversations tend to become dominated by a single partner as students attempt to earn maximum credit for a correct answer. We also present a comparison of two methods for studying conversation bias in peer instruction discourse that could be valuable to other researchers interested in studying peer discourse.

### 1. INTRODUCTION

Electronic classroom response systems (CRSs) have been in use in large college lectures for over three decades (Bruff 2007). Such systems are designed to provide instructors and students with immediate statistical analyses of student electronic responses to multiple-choice questions posed to the class by the instructor. Instructors may choose to include student CRS scores in their overall course grading scheme, and they have the option to adjust a grading incentive within the CRS grade book, which assigns 0–100% credit for incorrect responses to reward class participation.

In the cooperative learning technique known as peer instruction, learners are provided time to discuss ideas with seat partners before registering individual CRS responses (Mazur 1991). Much research has been done to study the impact of peer instruction on student learning and engagement (Beatty et al. 2006; Cortright, Collins, & DiCarlo 2005; Meltzer & Manivannan 2002). Some authors have encouraged the inclusion of CRS scores as part of student assessment to motivate participation (Duncan 2006; Len 2007). However, little research has been done on how different assessment strategies for the inclusion of CRS scores influence the type of student discourse that occurs in the technique.

In this study, we build on a previous study's findings (James 2006) that indicate that the assessment practices of instructors who use the peer instruction technique have a significant impact on the quality of peer discourse that occurs as a part of the technique. In our current study, we provide further support for this finding, comparing discourse data between two identical classes taught by the same instructor who used contrasting assessment strategies. We also demonstrate the effectiveness of a much simpler method for analyzing peer discourse that would be of interest to other researchers seeking to study peer discourse.

## **2. DESIGN/METHODS**

### **2.1 Sample and Design**

Our study was conducted over two semesters. In the first semester, two astronomy professors teaching large-enrollment introductory astronomy classes were recruited to use CRS to implement peer instruction in their classes. Each instructor was assigned a graduate assistant whose duties were to operate the CRS system on a daily basis and collect data for use in this study. Both instructors implemented the peer instruction technique using the same procedure; that is, they conducted a minimum of three CRS question sessions each class period and provided time for students to discuss ideas prior to entering individual electronic responses. Instructors created their own respective CRS questions for use in peer instruction.

Data on student discourse that occurred in response to CRS questions were collected during class periods near the beginning, middle, and end of each course. In the first semester, a random sample of 28 participants from Instructor A's class and 24 participants from Instructor B's class used audio recorders to tape their CRS conversations during these class periods. Additional data were collected from a random sample of 32 participants from Instructor A's second-semester class. Demographics for all three classes were similar and are provided in Table 1.

<b>Table 1. Student Demographics</b>			
	<b>Instructor A Semester 1</b>	<b>Instructor A Semester 2</b>	<b>Instructor B Semester 1</b>
Course Title	Intro to Astronomy	Intro to Astronomy	Life in the Universe
Class Size	180	178	84
Student Majors	89% Non-science majors	85% Non-science majors	84% Non-science majors
University Level	66% freshmen	61% freshmen	64% freshmen
Student Gender	56% male	52% male	56% male

The two instructors in this study initially had differing philosophies regarding how CRS scores were to be included in students' overall course grades. In Semester 1 of this study, Instructor A adopted a high-stakes approach in which the grading scheme motivated students to answer CRS questions correctly. Accordingly, incorrect responses earned only one third the credit earned by a correct response, and a student's overall CRS score counted for 12.5% of the overall course grade. In contrast, Instructor B adopted a low-stakes approach in Semester 1 in which student CRS scores were computed solely based on degree of participation. In this approach, a student's overall CRS score counted for 20% of overall course grade, and incorrect responses earned as much credit as correct responses.

In Semester 2, Instructor A changed to a low-stakes grading approach in which incorrect CRS responses earned 90% of the credit that a correct response would earn. The percent of overall course grade for CRS scores remained unchanged at 12.5%. The peer instruction technique administered in Instructor A's second-semester class was identical to that used in Semester 1. Semester 2 CRS questions used by Instructor A were also identical to those used in Semester 1.

## **2.2 Measuring Discourse Bias: Two Complementary Techniques**

Discourse bias here is defined as the difference between the fractional contributions to a conversation between partners. If, for example, a conversing pair had one partner contributing 80%, the other partner's contribution score would then be 20%, and the pair's discourse bias would be 60%. This quantity was used to reflect the relative balance between partners that existed in a conversation transcript.

Discourse bias was measured using two techniques. The first technique involved categorizing the content uttered by talking partners during their recorded CRS conversations. Each idea (or utterance) was coded according to 10 categories adapted from Kaartinen and Kumpulainen's (2002) techniques for discourse analysis: 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 other partner's idea; and expressing uncertainty. This technique was very time consuming and required careful training of coders to achieve a high interrater reliability correlation of  $r = 0.922$ . This content-based technique was initially implemented in the hope that the analyses of transcripts would reveal correlations between various idea codings and CRS question types. Analyses, however, did not reveal significant correlations between question type and idea categories. The limitations of this approach are discussed next.

A second technique to determine discourse bias adopted in the present study involved investigating the relative amount of student talk produced by each conversation partner during a peer instruction question session. In linguistic research, amount of talk has been studied using a variety of measures (James & Drakich 1993), including total number of words (McFayden 1997), amount of time spent talking, number of turns, and average number of words per turn (Edelsky 1993). Total number of words is perhaps the most reliable and easily replicable of these measures. Accordingly, in the present study, amount of talk was operationalized as the total number of words produced by a student during a peer discussion. The amount of talk of individual students in peer instruction sessions was captured using an specially designed computer program (see Note 1) that, by identifying speaker turn codes, splits original transcripts into separate transcripts for each individual speaker participating in peer discussion. Once individual transcripts were generated for individual students' talk, the number of words included in these "new" individual transcripts was counted using an additional computer program (see Note 2).

## **2.3 Measuring Partner Consensus in CRS Question Responses**

Student responses to CRS questions were electronically recorded and saved in each instructor's respective class database. At the end of the study, partner responses during recorded sessions were compiled, and the frequency of disagreement between partners was tallied.

# **3. FINDINGS**

## **3.1 The Influence of Grading Incentive on Peer Discourse**

As shown in Table 2, the word counting technique yielded similar bias scores to the idea coding technique, with a high overall correlation between the two techniques of determining discourse bias with  $r = 0.856$ . Further, biases calculated using the word counting technique tended to be somewhat higher than idea code tabulations. This may possibly result from the fact that different idea codes are often expressed in widely discrepant numbers of words. For example, it typically takes many more words to generate an idea code for "providing justification" than it does to "state agreement," but each idea code generates the same weight in the overall tabulation of idea count.

<b>Table 2. Discourse Bias and Partner Question Disagreement</b>			
	<b>Discourse Bias (Idea Count Technique)</b>	<b>Discourse Bias (Word Count Technique)</b>	<b>Partner Question Response Disagreement</b>
Instructor A, Semester 1 High-Stakes Grading	33.2% ( <i>SD</i> =30.1%)	39.8% ( <i>SD</i> = 30.7%)	7.6 %
Instructor A, Semester 2 Low-Stakes Grading	19.5% ( <i>SD</i> = 15.4)	26.6% ( <i>SD</i> = 19.2%)	17.1%
Instructor B, Semester 1 Low-Stakes Grading	14.6% ( <i>SD</i> = 10.7%)	21.5% ( <i>SD</i> = 15.4%)	36.8%

Results suggest that the assessment practices of instructors have a significant impact on the peer discourse that occurred as a part of the technique. As shown in Table 2, when the high-stakes instructor changed her assessment strategy to a low-stakes approach in the second semester, discourse bias measurements became significantly lower in contrast to her previous semester, with mean bias dropping from 33.2% to 19.5%, with  $F(1,58)=5.268, p = 0.025$ , based on the idea count technique. Analysis of variance (ANOVA) calculations using statistics obtained from the word count technique yielded similar results, with mean bias dropping from of 39.8% to 26.6%, with  $F(1,58)=4.220, p = 0.044$ . Discourse bias scores in the two low-stakes classrooms taught by different instructors were statistically indistinguishable. These findings suggest that when instructors adopt a low-stakes grading strategy for CRS, peer discourse becomes more balanced between partners.

The frequency of partner CRS question responses that reflected disagreement increased significantly when the high-stakes instructor adopted a low-stakes grading strategy in the second semester, with mean partner disagreement increasing from 7.6% to 17.1%, with  $F(1,195) = 4.065, p = 0.045$ . Partner disagreement in the other low-stakes classroom was also very high. These findings suggest that when instructors adopt a low-stakes grading strategy for CRS, peer discourse becomes less focused on the identification of correct responses, and students are more likely to register CRS responses that reflect their own ideas. It follows that question response statistics displayed by a CRS system after each question in a high-stakes classroom may exaggerate the degree of understanding that actually exists because partners tend to answer with the response identified by the more confident partner. Therefore, the ability of an instructor to make accurate pedagogical decisions based on student response feedback is obstructed when using a high-stakes grading approach. The significant difference between the levels of partner question response disagreement in the two low-stakes classrooms taught by different instructors may be a result of individual differences in instructor question-writing style.

## 3.2 Limitations

Pedagogical interventions by both instructors during peer dialogues confounded the researcher's ability to determine clear statistical correlations between question types and measures of student discourse. During CRS peer dialogues, both instructors monitored nearby student conversations and sometimes interrupted the peer discussions to provide additional details, clarifications, or points for consideration. Although these pedagogical interventions were certainly laudable, they challenged the researcher's ability to categorize particular questions because additional information was sometimes added verbally during class. Additionally, both instructors limited discussion session times based on the number of students entering electronic responses, sometimes prematurely truncating recorded student conversations.

## 4. CONCLUSIONS

The assessment practices of instructors using the peer instruction technique for large-enrollment college science courses have a significant impact on the peer discourse that occurs in response to this technique. Findings suggest that conversations in high-stakes classrooms tend to become dominated by a single partner as students attempt to earn maximum credit for a correct answer. Further, our findings suggest that talking partners in high-stakes classrooms are more likely to provide identical CRS responses compared with partners in low-stakes classrooms. This suggests that question response statistics in low-stakes classrooms may be more representative of the degree of understanding that actually exists in a class. And finally, the word count technique provided similar results to the idea count technique for measuring discourse bias and therefore provides future researchers with a much simpler way to reliably measure conversation contribution.

## NOTES

**Note 1:** We are grateful to Camilla Vasquez for allowing permission to use the "Camilla Speaker Splitter" program.

**Note 2:** MonoConc Pro is a powerful text searching program commonly used in the field of linguistics. It is available at <http://www.athel.com/mono.html#monopro>.

## Acknowledgment

We are grateful to Northern Arizona University e-Learning director Don Carter for his support of this study.

## References

- Beatty, I. D., Gerace, W. J., Loenard, W. J., & Dufresne, R. J. 2006, "Designing Effective Questions for Classroom Response System Teaching," *American Journal of Physics*, 74(1), 31.
- Bruff, D. 2007, "Classroom Response Systems." Retrieved January 6, 2008, from Vanderbilt University, Center for Teaching Web site:  
[http://www.vanderbilt.edu/cft/resources/teaching\\_resources/technology/crs.htm](http://www.vanderbilt.edu/cft/resources/teaching_resources/technology/crs.htm).

- Cortright, R. N., Collins, H. L., & DiCarlo, S. E. 2005, "Peer Instruction Enhanced Meaningful Learning: Ability to Solve Novel Problems," *Advances in Physiology Education*, 29(1-4), 107.
- Duncan, D. 2006, "Clickers: A New Teaching Aid with Exceptional Promise," *Astronomy Education Review*, 5(1), 70. <http://aer.noao.edu/cgi-bin/article.pl?id=194>.
- Edelsky, C. 1993, "Who's Got the Floor?" in *Gender and Conversational Interaction*, D. Tannen (Editor), New York: Oxford University Press, 189.
- James, D., & Drakich, J. 1993, "Understanding Gender Differences in Amount of Talk: A Critical Review of Research," in *Gender and Conversational Interaction*, D. Tannen (Editor), New York: Oxford University Press, 281.
- James, M. C. 2006, "The Effect of Grading Incentive on Student Discourse in Peer Instruction," *American Journal of Physics*, 74(8), 689.
- Kaartinen, S., & Kumpulainen, K. 2002, "Collaborative Inquiry and the Construction of Explanations in the Learning of Science," *Learning and Instruction*, 12, 189.
- Len, P. M. 2007, "Different Reward Structures to Motivate Student Interaction with Electronic Response Systems in Astronomy," *Astronomy Education Review*, 5(2), 5. <http://aer.noao.edu/cgi-bin/article.pl?id=209>.
- Mazur, E. 1991, *Peer Instruction: A User's Manual*, Upper Saddle River, NJ: Prentice Hall.
- McFayden, R. G. 1997, "The Relationship between Powerless Speech, Agentic Behavior, and Amount of Talk," *Journal of Social Psychology*, 137(4), 470.
- Meltzer, D. E., & Manivannan, K. 2002, "Transforming the Lecture-Hall Environment: The Fully Interactive Physics Lecture," *American Journal of Physics*, 70(6), 639.