AMERICAN JOURNAL of PHYSICS

Listening to student conversations during clicker questions: What you have not heard might surprise you!

Mark C. James and Shannon Willoughby

Citation: American Journal of Physics **79**, 123 (2011); doi: 10.1119/1.3488097
View online: http://dx.doi.org/10.1119/1.3488097
View Table of Contents: http://scitation.aip.org/content/aapt/journal/ajp/79/1?ver=pdfcov
Published by the American Association of Physics Teachers

Listening to student conversations during clicker questions: What you have not heard might surprise you!

Mark C. James
Department of Physics and Astronomy, Northern Arizona University, Flagstaff, Arizona 86011

Shannon Willoughby
Department of Physics, Montana State University, Bozeman, Montana 59717

(Received 10 August 2009; accepted 19 August 2010)

When instructors provide time for students to discuss their ideas in Peer Instruction, instructors minimally expect that the conversation partners will discuss their opinions relating to the physical attributes posed in a question and submit clicker responses that coincide with individual opinions. We defined conversations that met these two criteria as "standard conversations." In our study of 361 recorded Peer Instruction conversations from large introductory astronomy classrooms taught by experienced instructors, we found that 38% of student conversations were standard conversations. Of the remaining 62%, we identified three broad categories consisting of ten types of "nonstandard" conversations. The first category of conversations describes student ideas that were not reflected in any of the given multiple choice answers. The second category includes issues related to the interpretation of the statistical feedback provided by electronic classroom response systems. The third category describes common pitfalls experienced by students during conversations that led to unproductive interactions. Our analysis of nonstandard Peer Instruction conversations will be useful to practitioners and researchers seeking to improve the implementation of Peer Instruction.
[DOI: 10.1119/1.3488097]

I. INTRODUCTION

Clickers have been used in the classroom since the early 1970s[1] but only recently have their use become widespread. Early studies of electronic response systems reported several positive effects including increased student engagement and classroom attendance.[2] These same studies often indicated no observable difference in learning gains as a result of the inclusion of this technology. In contrast, studies focusing on the efficacy of clicker implementation schemes that incorporate other active learning techniques have reported positive learning gains. Mazur[3] found profound learning gains when using a classroom response system to facilitate small group discussions during lectures in a technique known as Peer Instruction. Mazur's results have been replicated in a number of contexts.[4–7] Turpen and Finkelstein studied Peer Instruction in actual practice and found large variations in how individual instructors implement the technique.[8]

When an instructor engages students in an active learning technique such as Peer Instruction, the instructor is not idle. Mazur notes that listening to student conversations allows him "to assess the mistakes being made and to hear how students who have the right answer explain their reasoning."[3] Duncan notes that listening to student conversations can provide unexpected insights into student ways of thinking.[9]

In a typical Peer Instruction session in a large enrollment introductory science classroom, there may be over 100 conversations going on at any one time. Although the tactic of mingling among students to listen to their conversations provides an instructor with valuable real-time feedback, the samples of student ideas obtained by this technique come from only a small number of groups that can be easily visited by the instructor during a given session. The sampling is further limited by the fact that as instructors seek to sample more conversations from the class, they must listen to smaller parts of individual conversations. Clearly, there are aspects of student conversations that are not available to instructors through the technique of real-time conversation sampling. Our study seeks to fill this gap by a systematic study of entire Peer Instruction conversations recorded from a random sampling of conversation partners in five large enrollment introductory astronomy classrooms.

The idea for our study originated as a result of mentoring colleagues participating in a university-wide initiative to popularize the use of clicker systems in large enrollment introductory classrooms. We conducted presentations, workshops, one-on-one meetings with faculty, and classroom visits with instructors from various disciplines seeking to take advantage of clicker technology to facilitate the pedagogical technique of Peer Instruction in large enrollment classrooms.

As instructors learning to use Peer Instruction developed clicker questions to stimulate discussion, they frequently designed multiple choice questions that included distracters highlighting common student misconceptions the instructors had observed in their often extensive teaching experience. In our conversations with these instructors, we found that they invariably imagined idealized productive conversations wherein students would systematically discuss various question alternatives in light of their own prior knowledge and experience. Beatty *et al.*[10] described these types of conversations in their pivotal work on designing clicker questions: "Typically, students within a group will argue their various opinions and intuitions, work out a solution if required, and continue discussing and elaborating until satisfied with their answer."

While working with faculty implementing clickers in their classrooms, we also collected recordings of Peer Instruction conversations as part of our studies on the effect of grading incentive on student discourse.[11–13] As we analyzed transcripts of these recordings, it became apparent that a large fraction of the student conversations diverged significantly from the idealized conversations that these instructors had envisioned as they constructed their clicker questions.

The primary goal of our current study is to provide practitioners with insights into the types of student conversations in their classrooms that depart from the idealized productive conversations instructors often envisage. Our hope is that practitioners and researchers will use this information to help develop more efficacious techniques that take greater advantage of the diversity of student ideas that exist in their classrooms and help students experience more meaningful peer conversations.

II. SAMPLE/METHODS

Data were collected from four sections of standard freshman level introductory astronomy courses for nonscience majors and one section of introductory astronomy that utilized a nonstandard multidisciplinary approach to study the special issues concerning space travel and the possibility of extraterrestrial life. Enrollment in the standard introductory astronomy courses ranged from 178 to 200 students per section, and enrollment in the multidisciplinary course was 84 students. Student demographics in all courses were similar with nearly even gender distributions, 84%–89% nonscience majors, and 61%–66% first-year enrollment.

Three astronomy professors at two midsized state universities in the western U.S. participated in our study. Instructor 1 taught two sections of Introduction to Astronomy and had seven years of experience teaching at the university level with a Ph.D. in theoretical condensed matter physics. Instructor 1 was a coauthor of one other study on clickers and coauthor of the current study.[11] She was the only instructor in our study with previous experience using clickers. Instructor 2 taught two sections of Introduction to Astronomy and has a Ph.D. in planetary sciences. Instructor 2 earned four awards for excellence in undergraduate teaching during her twenty-four years of experience teaching at the university level. Instructor 3 taught one section of Life in the Universe and has a Ph.D. in planetary sciences. During his seven years of experience teaching at the university level, he earned a distinguished teaching award, authored a popular book on astrobiology, and is currently writing a textbook for astrobiology.

The instructors in our study utilized a variant of the Peer Instruction technique whereby each clicker session began with the instructor posing a multiple choice question to the class using a large-screen projector. Students were then provided time to discuss their ideas with their seat neighbors before they entered their individual responses to questions using hand-held clickers. At the end of each question session, the instructor displayed the student response statistics and discussed the various responses. This technique was identical to Mazur's technique[3] except that the step where students silently consider and enter individual responses before engaging in peer discussions was omitted due to our instructors' concerns regarding the additional class time required. Similarly, none of the faculty in Turpen and Finkelstein's study[8] of the implementation of Peer Instruction utilized the individual silent response step.

The instructors adapted their clicker questions to meet various individual pedagogical goals. Instructors were interviewed after classroom data collection was complete. Table I describes the pedagogical motivations of instructors for nine typical clicker questions that were posed during this study. Most of the clicker questions were developed by the instructors, but questions were occasionally adapted from other sources. Instructors' motivations for designing questions ranged from stimulating critical thinking skills to explicitly encouraging students to complete reading assignments. The inclusion of factual-recall type questions in addition to higher level conceptual oriented questions for Peer Instruction differs from the technique originally described by Mazur.[3] The wide spectrum of question types used by the instructors in our study was consistent with the various types of questions observed in other natural settings.[8]

Tape recordings of student discourse during clicker sessions were collected during several class periods throughout each course. Each recording was transcribed for subsequent analysis. The subsets of students in each class who participated in the study were selected at random from the classes and were provided a chance to participate in a raffle held at the end of each course as an incentive to take part in the taping. The fraction of female students in our sample was 50.1%. Information relating to the ethnic make up of students was not obtained. In all, 147 students participated in the 45 clicker question sessions, and 361 conversations were transcribed.

To determine differences in student conversations that might result from different course grading incentives, Instructor 1 and Instructor 2 each taught two sections of the same course with contrasting grading schemes for clicker question responses. In Instructor 1's high stakes grading section, only correct clicker responses earned credit in contrast to the low stakes grading section where incorrect and correct responses earned the same credit. Clicker scores accounted for 4% of the student's overall course grade. In Instructor 2's high stakes grading section, incorrect responses earned one-third as much credit as a correct response; incorrect and correct responses were credited equally in the low stakes grading section. Clicker scores accounted for 12.5% of the student's overall course grade in Instructor 2's sections. Instructor 3 adopted a low stakes grading scheme in which incorrect and correct responses were counted equally, and clicker responses accounted for 20% of the overall course grade.

Our goal in the qualitative analysis of the conversation transcripts was to catalog the types of conversations that deviated from the idealized expectations instructors had shared with us in our work mentoring faculty. The first step in our analysis was to define the minimum criteria for a "standard conversation" that would satisfy the minimum expectations of these instructors. We defined a standard conversation as a conversation where conversation partners discussed aspects of at least one multiple choice alternative provided in a clicker question, and individual clicker responses were representative of ideas that individuals had articulated. We then used the qualitative technique of constant comparison[14] to develop conversation types as we analyzed transcripts that did not meet these two criteria for a standard conversation. As conversation types emerged from this analysis, transcripts were categorized, and these categories were again studied for consistencies and differences to further classify conversation types. This process was repeated over many cycles, and we

Table I. Questions and instructor pedagogical rationales. Conversation examples resulting from 9 of the 45 questions observed in this study were highlighted to illustrate the various conversation types focused on in this paper. Correct answers are in underlined. Instructor's pedagogical rationales for including each question are in italics.

(1) If everything in the solar system is moving around, why does the Perseid meteor shower repeat around August 11 every year? (a) Because the Earth does not move relative to the meteor particles. (b) Because the Earth in its orbit intersects the same swarm of meteor particles at the same time each year. (c) Because the Perseids are not in space at all, but in the upper layers of the Earth's atmosphere. (d) Because August 11 is when the Sun's rays are the warmest and thus tend to move the meteor particles toward us at the fastest speed.

Question 1 was obtained by Instructor 2 from a test bank included with the textbook's ancillary materials, and had been included in previous semesters as a test question. This question was posed to determine how well students understood that meteor showers are caused by Earth's motion through material left behind by a comet. The pictures demonstrating this concept in the course textbook show comet debris as stationary as Earth passes through. Although students learn that everything is moving in the solar system, that concept does not always "sink in," possibly due to the static diagrams in their text. This question emphasizes the motions in the solar system while also testing their understanding about the cause of meteor showers. The students thus are tested on two concepts in one question.

(2) How much of the mass of the Solar System is contained within the *Sun*? (a) 63%. (b) 25%. (c) 75%. (d) 99%. (e) Cannot conclude.

Instructor 1 intended this question as a simple recall level question to review a point from a previous lecture regarding the relative mass from the solar system that is contained in the Sun.

(3) How much more massive is Jupiter than the Earth? (a) 18. (b) 50. (c) 318. (d) 502. (e) 5047.

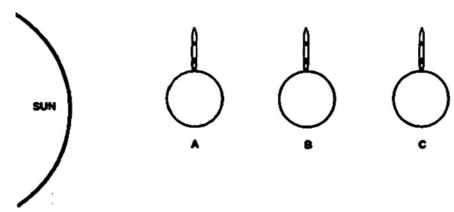
Following Mazur's suggestions, Instructor 1 often used her clicker system to motivate students to complete reading assignments prior to coming to class. Because the primary goal of such questions was to reward students for completing reading assignments, questions such as this one were designed to be simple to answer if a student has completed the assignment.

(4) Why is the Earth hot inside? (a) Residual heat trapped from formation. (b) A tiny fusion core of Hydrogen at the center. (c) Heating by the same dynamo that makes Earth's magnetic field. (d) Heat due to contraction. (e) Heating from radioactive elements.

The Earth's internal energy is important to astrobiology as a driver for plate tectonics and the carbon cycle. It is also an indicator of Earth's age. The fact that so much thermal energy could be generated in Earth's interior by native radioactive materials is intuitively unexpected. Instructor 3 wrote this comprehensive level question to pose students with contrasting answers, some of which are correct for other bodies, but not Earth.

(5) From which planet (all of equal mass) would it be easiest for a rocket to take off from? (a) Closest to sun. (b) Next one out. (c) Third from Sun. (d) Same for all three. (e) Cannot conclude.

Instructor 1 took this question verbatim from a concept inventory on gravity (Ref. 22). This question was posed along with two others from Ref. 22 during a lecture on gravity to probe student's understanding of gravity.



(6) Why do stars in the halo of our Galaxy have little of the common elements such as carbon, nitrogen, and oxygen? (a) These elements have been used up by the stars. (b) C, N, and O are biological elements and there is no life out there to make them. (c) The halo stars are old. They formed before those elements were made. (d) Making C, N, and O requires massive stars and there are no massive stars in the halo.

A primary focus of Instructor 2's class is the development of students' logical thinking skills. Students learn from the text and in class that carbon, nitrogen, and oxygen are created in stars more massive than the Sun during the star's dying stages. They also learn in class that these stars explode as supernova and distribute these elements into the interstellar medium. Stars which form out of this enriched material will therefore show higher levels of C, N, and O than stars that formed from gas/dust not containing these elements. Stars with higher concentrations of C, N, and O therefore tend to be younger than stars depleted in these elements. The students also learn in class that the halo stars are the first stars to form in a galaxy. The instructor expected students to use logical thinking skills to put these various concepts together to arrive at the correct response.

(7) Name one difference between green light and x-rays. (a) Speed at which they travel. (b) Wavelength. (c) X-rays are not light. (d) X-rays do not spread out. (e) Green light does not spread out.

Question 7 was adapted by Instructor 1 from a concept inventory of light and spectroscopy (Ref. 23) to gauge students' understanding of material that had just been presented in the lecture.

(8) Which will fall faster in a vacuum? (a) A feather. (b) A ball. (c) They will fall at the same rate. (d) Cannot conclude.

Instructor 1 was interested in gaining insight into students' prior knowledge of gravity. Students were asked to predict the outcome of this demonstration before it was performed by the instructor.

(9) Earth's current atmosphere is composed primarily of (a) carbon dioxide and nitrogen, (b) water vapor and oxygen, (c) hydrogen and helium, (d) nitrogen and oxygen.

This question was designed by Instructor 2 as a reading test as suggested by Mazur (Ref. 3).

ultimately observed three broad categories of conversations that were further classified into ten types. Each nonstandard conversation was then coded according to these conversation types to obtain descriptive statistics. Fifty seven conversations (15% overall) were coded by both researchers, with an inter-rater reliability of 89.3% for our coding procedure.

III. FINDINGS

Of the 361 conversations studied, 136 were identified as standard conversations in which partners discussed aspects of at least one multiple choice alternative and clicker responses were representative of ideas that individuals had articulated. The remaining 225 conversations were classified according to three broad categories to highlight different aspects of these conversations. Some conversations were consistent with more than one category. See Table II for a summary of all conversation types.

A. Unanticipated student ideas

We found that 12.5% of the student conversations included instances where multiple choice alternatives provided with the clicker questions did not encompass the entire spectrum of incorrect student ideas that were discussed. We highlight four types of peer discussions that are representative of this category.

1. Unanticipated student ideas regarding directly related content knowledge

In 3.9% of the conversations, students discussed incorrect ideas directly related to the content focus of a question that were not represented in one of the multiple choice answers. Student responses to Question 1 in Table I illustrate this conversation type. Responses (a), (c), and (d) reflect common misconceptions that were familiar to the instructor. Of the 13 student groups whose conversations were recorded on this day, three groups had elaborate discussions relating to the same incorrect idea not anticipated by the instructor in any of the question alternatives. In the following conversation excerpt, Bob struggles to understand how Earth could intersect the remnants of a comet in the same place in space every year because the comet fragments and Earth would surely have different orbital periods. During their conversation, Ted is able to successfully explain to Bob that comet fragments are distributed throughout the comet's orbital path like a loop of string, so Earth must necessarily intersect different fragments during each yearly meteor shower.

> Bob: "The Earth is in an orbit. Why don't meteoroids move around?"
>
> Ted: "Because it's the comet's orbit. So, wouldn't it just be a big circle moving around in the same place?"
>
> Bob: "It probably wouldn't be the same length as the Earth's, right? So if one year it started here, you know, they could be on a ten year orbit."
>
> Ted: "Well it doesn't matter. It is just a constant string."
>
> Bob: "Hopefully, all the way around so it's not the same ones. Yeah, yeah."
>
> Ted: "We're just hitting the string of particles."

2. Unanticipated student ideas regarding near related content knowledge

In 1.1% of conversations, students discussed incorrect ideas nearly related to the content focus of a question that were not represented in one of the multiple choice answer responses. In such cases, student ideas sometimes led them to question concepts that the instructor had implicitly assumed students understood when creating the question. In Question 2 from Table I on material covered in a previous lecture, students were asked what percentage of the mass of the solar system is contained within the Sun. In creating this question, the instructor implicitly assumed that students understood what types of objects compose the Solar System. In the following excerpt, it was found that this review question caused Christina to reveal her notion that stars could be found within the solar system.

> Shawn: "99%."
>
> Christina: "I'd say 75."
>
> Shawn: "99. I remember that from lecture."
>
> Christina: "I say 75, because there's a lot of, there's what, stars and planets and stuff."
>
> Shawn: "Our solar system. Not our *galaxy*… our *solar* system."

3. Unanticipated student ideas regarding fundamental science content knowledge

As instructors create questions, they necessarily assume students share some common level of fundamental science content knowledge. Yet many students come to introductory classes with fragmented and incomplete understandings of basic scientific concepts. In 4.2% of the conversations, students articulated incorrect understandings of fundamental science content knowledge that interfered with their ability to successfully consider specific aspects of a clicker question. In the following example, the instructor intended Question 3 from Table I as a simple recall question to review material from a previous lecture, with the implicit assumption that students understood the concept of mass. In response to this question, Cody expressed an incorrect idea regarding the mass contained in solid and gas phases of matter. This conversation revealed Cody's difficulty with the concept of mass and allowed his conversation partner to highlight the scientifically accepted understanding.

> Cody: "When they say mass, they don't mean gas because gas would not have any mass because it's not solid."
>
> Eric: "Mass as in like material, like atoms or something."
>
> Cody: "Gas has fewer material."
>
> Eric: "No, it just takes up more space because it's in its gaseous form."
>
> Cody: "Right, so like it has less material for like

Table II. Nonstandard conversation types.

Conversation category	Conversation type	Description
(A) Unanticipated student ideas	(A1) Unanticipated student ideas regarding directly related content knowledge	A conversation where students discuss an incorrect idea directly related to the instructor's intended content that was not reflected in any of the multiple choice answer alternatives
	(A2) Unanticipated student ideas regarding near related content knowledge	A conversation where students discuss an incorrect idea regarding content knowledge not directly related to the instructor's intended content
	(A3) Unanticipated student ideas regarding fundamental science content knowledge	A conversation that reveals a gap in fundamental science content knowledge that is prerequisite to a consideration of the clicker question
	(A4) Trying to relate disparate material from previous lectures	A conversation where students naively seek to draw consistency with previous ideas from the class
(B) Statistical feedback misrepresenting student understanding	(B1) Using extraneous cues to arrive at a response	A conversation where students arrive at a response based on guessing strategies, religious views, or incorrect information not represented in any answer alternative
	(B2) Responding with another student's answer preference	A conversation where a student responds with a partner's answer preference that is not consistent with the student's answer preference articulated in the conversation
(C) Conversation pitfalls	(C1) Not articulating physically based rationales when making choices	A conversation where answer alternatives are considered without providing any physically based rationale
	(C2) Answer appears self-evident	A conversation that is prematurely concluded when a partner indicates a particular answer is obviously correct
	(C3) Passivity in deference to another student	A conversation where only one partner articulates ideas relating to physical aspects of a question
	(C4) Students unable to initiate a conversation	A conversation where students express inability to initiate a discussion of physical attributes

whatever you want to use to measure the space."

In some cases, gaps in fundamental science content knowledge led students to doubt an implicit assumption within a question. Question 4 from Table I was used by the instructor at the beginning of a new unit on planetary geology to probe student's existing knowledge of the mechanism responsible for maintaining Earth's high internal temperature. In addition to the correct response that refers to radioactive decay, the instructor provided four plausible distracters based on common misconceptions he had identified while previously teaching the course. In the following excerpt, a gap in Alan's fundamental science content knowledge leads him to doubt whether any of the answer alternatives could be correct.

Alan: "Why is the Earth hot inside?"

Laura: "Heat due to contraction."

Alan: "But, why do you think the earth is *HOT* inside?"

> Laura: "Oh, well, liquid hot magma."
>
> Alan: "Liquid hot magma? I don't know why the Earth would be hot inside because… I mean when you go to like the frozen tundras and that sort of thing… it's permanent… it's like permafrost."
>
> Laura: "Yeah."
>
> Alan: "And it takes on whatever temperature is outside. So how could the heat penetrate that deep?"
>
> Laura: "But the Earth's crust is a certain amount of miles deep, which is crusted over."

4. Trying to relate disparate material from previous lectures

In 3.3% of the observed conversations, students naively attempted to relate concepts from the current question with unrelated ideas covered previously in the course. In such cases, it is possible for students to generate new incorrect understandings of the material. In the following conversation relating to gravity in response to Question 5 from Table I, students were asked to choose the planet from which it would be easiest for a rocket to take off from. This question was presented soon after a lecture on relativity and black holes, complicating the concept of gravity for one particular student. Zeke appears to have a grasp of the concepts necessary to deduce the correct answer (that proximity to the Sun does not matter, merely the mass and size of the planet), but Chelsea is struggling with possible relativistic effects influencing planets closer to the Sun.

> Zeke: "You're not trying to escape the gravity of the Sun; you are trying to escape the gravity of the planet."
>
> Chelsea: "Yeah, but don't you remember when we talked about the Sun being weird and time slowing down and all that weird stuff? Does that matter? I don't know if that would be an effect."
>
> Zeke: "If time slowed down, then ultimately you'd burn fuel faster, but you'd also… No you'd burn fuel slower, but you'd move slower… They are all sort of space and time."

B. Statistical feedback misrepresenting student understanding

One of the powerful features of using the Peer Instruction technique in concert with an electronic feedback system is the ability for instructors to gain immediate feedback on student understanding via question statistics that are displayed as students enter their responses. The ability of instructors to obtain this type of real-time feedback in a large lecture setting is used as a selling point by the manufacturers of clicker systems.[15] However, in 26% of the observed conversations, we found that the statistical feedback generated by these systems misrepresented the nature of the existing student understanding. In the following, we present two conversation types that highlight issues related to the interpretation of statistical feedback generated by clicker systems.

1. Using extraneous cues to arrive at a response

Students utilized cues that were tangential or irrelevant to the concepts highlighted in answer alternatives to arrive at a response in 21.9% of the conversations. Strategies in this conversation type included searching by rote for terms the instructor had uttered recently, searching for cues in the question's phrasing, recalling answers to seemingly related test questions, blindly guessing, or referring to religious views to select a response. Conversations were also included in this category when students chose a response based on an unresolved incorrect idea from a type A conversation that was not represented in one of the possible answers.

Instructor 2 expected that students choosing the correct response for Question 6 in Table I would be required to use sophisticated reasoning skills to put together three ideas previously covered in the course: The Big Bang explosion produced only light elements, stars in globular clusters formed shortly after the Big Bang, and all heavy elements present in the universe are created through nuclear fusion in stars. Of the 11 student groups whose conversations were recorded on this day, eight groups answered correctly with no disagreement between partners. Yet, only one of the eight groups that gave correct responses put the three critical concepts together in a coherent argument to obtain the correct response. The remaining seven groups obtained the correct answer using extraneous cues. In the following excerpt, two answer alternatives are eliminated by Gary based on a single statement of fact made by the instructor that appeared to correspond to a word that occurred in two of the four alternatives. Dave had a sense that the age of the stars in the halo is key, but the students struggled unsuccessfully to resolve an apparent inconsistency introduced by Gary's incorrect idea about the origin of heavy elements.

> Gary: "She was saying that the populations of halo stars are old."
>
> Dave: "Well, just because they are old doesn't mean that they form…"
>
> Gary: "The reason we have carbon, nitrogen, and oxygen is because our Sun is young."
>
> Dave: "Oh, OK… So those older stars wouldn't have it?"
>
> Gary: "No. They won't have it because they are in their dying stages and lose things like carbon."

2. Responding with another student's answer preference

In 4.2% of the transcripts, we found that students entered responses that did not correspond to their stated answer preference in deference to another student who appeared to have greater confidence. After a brief introduction to the nature of light, the instructor posed Question 7 from Table I regarding the differences between green light and x-rays. Although the instructor had intended this question to be a low level factual review of lecture material, the question sparked rich discussions in most groups that were recorded. 96% of the students responded with the correct answer. A simple interpretation of the question statistics would have led the instructor to believe that the vast majority of students in class understood that visible light and x-rays were composed of the same type of radiation but with different wavelengths. Yet, it was clear from the conversation transcripts that students were highly confused about the nature of radiation. The following excerpt was typical of the substantive discussions that took place in response to this question. Even though the issues raised in this discussion remained unresolved at the end of the session, every member of the group ultimately answered with the same correct response suggested by Jordan.

Jordan: "Green light, is that like night vision?"

Raquel: "Well, they travel at the same speed."

Jordan: "Because all light travels at *c*. Wavelength, maybe."

Dan: "X-rays don't spread out."

Jordan: "X-rays don't? Well, yeah, because they are just direct; it's a direct line."

Dan: "Um, but does green light spread out?"

Jordan: "Light does not spread out. It's going to be speed."

Raquel: "No, light does spread out."

Jordan: "Well, speed is constant between all forms of light."

Raquel: "That's what she said today in class; she showed the diagram of light spreading out. Does x-ray?"

Dan: "I don't know, but isn't the wavelength different?"

Jordan: "Yeah, it's like this. [Presumably gesturing.] So it would be wavelength."

C. Conversation pitfalls

Conversations were unproductive in 37.7% of the transcripts. The four types found in this category identify common characteristics of these conversations.

1. Not articulating physically based rationales when making choices

A technique student partners often used for determining a correct response was to begin their discussions by eliminating responses that appeared unlikely. This process often elicited substantive discussions when students provided some type of physically based rationale as a basis for eliminating alternatives. Yet, in 21.9% of the transcripts, partners discussed their relative certainty regarding various question alternatives without mentioning any physical attributes relating to the question. The following excerpt, typical of such exchanges, could have been created in response to any question because there is no discussion of physically based rationales.

Bill: "What do you think, (e)?"

Karen: "I don't know for sure… Do you think its (e)?"

Bill: "Yeah, I think that it's (e). What do you think?"

Karen: "It's not (c)… (b) just sounds weird too…, (e)?"

Bill: "Yeah, (e) sounds good … We're done."

2. Answer appears self-evident

In 8.9% of the conversations, one or more students indicated that a correct response seemed immediately apparent and conversations were prematurely concluded. Deeply held misconceptions often appear to students as commonsense facts whose correctness is beyond question. In the following conversation in response to Question 8, students quickly agree on a correct response to a question without examining the basis for any of their ideas related to the question. Although the classroom's overall correct response frequency was 90% for this question, subsequent probing by the instructor found that most students who responded correctly could not articulate a coherent rationale for their choice.

Jake: "They are obviously going to fall at the same rate."

Chris: "Yeah, based on what we have seen."

Jake: "Because gravity is going to be affecting both of them…"

Chris: "Yeah…"

Jake: "At the same rate."

3. Passivity in deference to another student

In 5.3% of the conversations, one student provided a physically based rationale to support an answer choice that was then accepted by partners without question. In this short exchange in response to Question 5 from Table I, Tim provided a rationale for choosing an incorrect answer. Chase registered the same incorrect response, but it is unclear whether he was swayed by Tim's argument, or if he possessed his own rationale for accepting the incorrect answer.

Tim: "It's A, because it's little. And it's close to the sun, and everybody knows that planets that are closer to the sun, rockets are easier to take off of."

Chase: "Yep, that's the one."

4. Students unable to initiate a conversation

In 1.7% of the transcripts, students did not initiate discussions regarding the question during the time allotted for peer discussion. In such cases, students sometimes spent their discussion time lamenting their lack of knowledge rather than attempting to examine ideas that may be related to the question.

D. Factual-recall type questions and conversation type

Of the 45 questions that were included in this study, 15 were designed as low level factual-recall type questions as a tool to motivate students to complete reading assignments or to review an important fact. Yet, the fraction of type A conversations where students generated unanticipated ideas in response to low level questions (13.6%) was statistically indistinguishable from the fraction of type A conversations resulting from higher level questions (11.9%), with $F(1359) = 0.191$ and $p = 0.662$.[16] The students in the following discussion of Question 9 from Table I carefully consider each alternative to this factual type question and, in so doing, reveal an incorrect idea about the ability for water vapor in air to exist in air.

> Julie: "Before we answer, let's look at all the common elements: It does have carbon dioxide in it, but we can't live in carbon dioxide and nitrogen."
>
> Janet: "Correct."
>
> Julie: "We can't live in water vapor and oxygen because we would drown. That's what I think. Hydrogen and helium? Without oxygen you're going to die. So, it has to be nitrogen and oxygen because the plants use nitrogen."

There also was no statistical difference in the overall fraction of standard conversations resulting from factual-recall type questions (34.7%) compared to the fraction of standard conversations resulting from higher level questions (39.1%), with $F(1359) = 0.637$ and $p = 0.425$. It might be expected that if there were any conversation types that would be affected by factual-recall type questions, it would be type C2 conversations where students articulated that a particular response was obviously correct. Yet, the fraction of C2 conversations resulting from factual-recall type questions (8.5%) was statistically indistinguishable from the fraction of C2 conversations resulting from higher level questions (9.1%), with $F(1359) = 0.033$ and $p = 0.856$. Comparisons between conversation codings for factual-recall type questions and higher level questions showed no statistical differences in any conversation type.

E. Grading incentive and conversation type

Instructors 1 and 2 both taught a high stakes section and low stakes section of the same course. It was found that a greater fraction of students in the high stakes classrooms either became passive in a conversation in deference to another student (type C3) or responded with another student's preference while indicating a different preference during the conversation (type B2). The fraction of Instructor 1's high stakes conversations in types C3 or B2 was 20%, in contrast to only 6.7% of Instructor 1's low stakes conversations. Results from Instructor 2's classes were similar, with 17.3% of high stakes conversations in types C3 and B2 compared to 7.0% of low stakes conversations. Although this difference between high and low stakes sections was statistically significant in Instructor 2's classes with $F(1,216) = 5.668$ and $p = 0.018$, the result in Instructor 1's class was not with $F(1,63) = 2.423$ and $p = 0.125$. Comparisons of high and low stakes classrooms for other conversation types did not show a statistical significance. Instructor 3 taught one course section with a low stakes grading incentive and it was found that 2.6% of conversations in this class were classified as C3 or B2 type conversations.

IV. DISCUSSION AND RECOMMENDATIONS

The instructors in our study designed clicker questions for use in Peer Instruction to meet a wide spectrum of pedagogical goals. These instructors also used the real-time statistical information provided by their clicker systems to focus their instruction on student needs. However, in 62% of the recorded conversations, students either discussed incorrect ideas not anticipated in the instructor's multiple choice alternatives or submitted clicker responses that were inconsistent with ideas that they had been articulated in their discussions.

Our first category of nonstandard conversation focused on conversations where students were engaged in discussing ideas that were not represented in the multiple choice alternatives provided by the instructor. In conversation type A1, students clearly understood the nature of the questions but expressed incorrect understandings of the subject matter that were not anticipated by the instructor. If these were the only type of unanticipated ideas discussed by students, we might conclude that the classroom technique could be perfected by simply including additional question alternatives to reflect a more exhaustive range of student ideas. However, such an approach assumes that all students possess the requisite level of knowledge that is implicitly assumed in the context of a clicker question. In conversation types A2 and A3, learners struggled with clicker questions because they lacked foundational content knowledge on which the question was based. And in conversation type A4, students struggled to search for meaning as they considered unrelated concepts from earlier parts of the course.

Even when clicker questions seemed incomplete or ill-posed to students, the discussions that often ensued demonstrated the potential of the Peer Instruction technique to facilitate a more active learning environment in which students cooperate to take steps to critically assess their understanding. In some cases, such as the sample conversation for type A3, we observed that the Peer Instruction technique enabled students to identify and begin to remediate incorrect understandings that might have otherwise remained concealed in a more passive learning environment. In other cases, such as the sample conversation in B2, we observed students struggling with incongruities within their conceptual frameworks that were too difficult to resolve in the short time allotted for discussion.

Beatty *et al.* emphasized the need for instructors to engage in a continuous process of reforming questions based on formative feedback from Peer Instruction sessions to generate higher quality dialogs that achieve the explicit goals for a course.[10] The instructors in our study rarely became aware of the unanticipated ideas that were being discussed in students' private conversations. Instructors wishing to enrich their lectures by addressing the diversity of ideas articulated in peer discussions might consider adopting explicit classroom protocols to encourage students to report divergent ideas that are elicited by peer conversations. This encouragement could be accomplished by asking students to share divergent ideas in a brief whole-class discussion at the end of each clicker session. Instructors might also provide a brief time after each question for students to provide some type of written feedback that could be collected after class.

Our analysis of student conversations revealed that clicker statistics often provided incomplete or misleading feedback to instructors. Type B1 conversations highlighted cases where students made answer choices based on cues that were tangential or irrelevant to a particular question, and type B2 conversations highlighted cases where students responded with another student's answer preference. Instructors wishing to utilize clicker statistics to better gauge the spectrum of student ideas in their classrooms might consider adding a none-of-the-above multiple choice alternative as recommended in Ref. 10 or ask students to use clickers to rate their degree of confidence in their response as recommended by Mazur.[3]

The example in response to Question 6 shows that students can sometimes choose a correct response to complicated conceptual questions without possessing a coherent understanding of the interrelated ideas the instructor had imagined when constructing the question. Incorrect responses to such a question do little to indicate to instructors which aspect(s) of the question is (are) causing difficulties. Instructors wishing to obtain a more detailed portrait of student understandings might consider giving a series of questions, each focusing on subsequent links in the conceptual chain. This technique is similar to that advocated by Meltzer.[5]

The contrast we observed between high stakes and low stakes classrooms suggests that if instructors adopt grading schemes where correct clicker responses are valued more highly than incorrect responses, students are more likely to either become passive in a conversation or select answers advocated by a partner that do not agree with their own ideas. Instructors who wish to improve the validity of clicker statistics and reduce the number of passive students in conversations could consider adopting a grading scheme where all clicker responses earn credit without regard to the correctness of the response. Instructors who wish to additionally utilize their classroom clicker systems as evaluation tools for scoring quizzes and tests can do so without influencing the quality of Peer Instruction clicker sessions by initiating separate evaluation-type clicker sessions.

All the students who participated in our study possessed highly developed conversational skills. Yet, we found that more than one-third of the conversations did not include any meaningful exchange of student ideas. In conversation type C1, students often conducted detailed assessments of their relative levels of certainty regarding various question alternatives without ever alluding to any physically based rationale. Such discourses cannot realize the optimal pedagogical impact of Peer Instruction. In type C2 conversations, the exchange of ideas was thwarted as students prematurely came to a consensus regarding a particular response that seemed obvious. Conversation types C3 and C4 illustrated cases where students either remained passive during a conversation or were unable to initiate a conversation.

One-third of the questions in this study were designed by instructors to test factual-recall type information. There were no observed statistical associations between this type of question and any of the conversation types that were cataloged. Students employed the same set of answer finding skills to all question types, regardless of the instructor's intended question level. An inherent strength of the Peer Instruction technique was certainly demonstrated when students were found to generate meaningful examinations of their own understanding in response to even low level factual-recall type questions. In contrast, when students approached high level conceptual type questions as if they were factual-recall type questions, as sometimes occurred in conversation types B and C, the Peer Instruction technique failed to realize its potential.

For peer discussions to realize their maximum pedagogical effect, students must critically assess the conceptual foundations that support their views as a fundamental part of every conversation. These skills form the basis for what is known as critical thinking and metacognition.[17] Instructors who wish to encourage their students to engage in dialogs that contain this sort of self-analysis could consider providing students with a short list of conversation behaviors that promote more meaningful discussions. Our proposed list of four behaviors to guide student interactions is adapted from similar lists that have proven effective in other contexts.[18,19]

(1) Assess aspects of every answer choice even if they seem incorrect.
(2) Generate your own answers when none of the given answers are consistent with your own ideas.
(3) Make note of new questions and ideas that arise as alternatives are considered.
(4) If you find a question is confusing and don't know how to begin your conversation, ask for help from your instructor or students nearby.

Because of the wide variability in student preparation in introductory science classrooms, students may perceive even the most carefully crafted question as ill-posed.[20] Novice learners are often challenged to reason in circumstances where their prior knowledge is loosely organized and/or incomplete, and they do not possess a problem-specific schema that enables them to systematically search for solutions.[21] A great strength of the Peer Instruction technique is that it allows students to address course material on their level. Mazur wrote that students can be more convincing than their professor,[3] which implies that student conversations can increase the amount of learning that takes place during a given lecture. We agree and add that the instructor also has much to learn from studying unexpected aspects of how students interact with one another when discussing clicker questions.

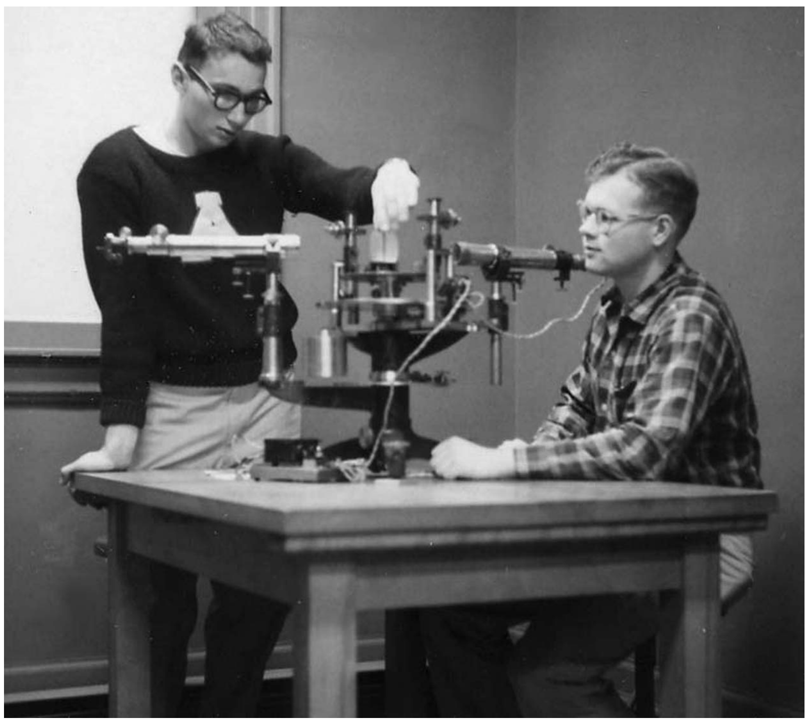
Further study is needed to establish how the various classroom norms and practices described by Turpen and Finkelstein[8] affect the frequencies of nonstandard conversation types. A study Peer Instruction conversation types in introductory physics classrooms where the student population is mostly science majors would also be of interest.

ACKNOWLEDGMENTS

The authors thank NAU e-Learning Director Don Carter and Assessment Coordinator Sue Pieper for their support of this study. The authors would also like to thank the reviewers who provided numerous insightful suggestions that helped the authors improve their manuscript.

[1] E. Judson, "Learning from past and present: Electronic response systems in college lecture halls," J. Computers Mathematics Science Teaching **21** (2), 167–181 (2002).
[2] J. Stowell and J. Nelson, "Benefits of electronic audience response systems on student participation, learning and emotion," Teaching of Psychology **34** (4), 253–258 (2007).
[3] E. Mazur, *Peer Instruction: A User's Manual* (Prentice-Hall, Upper Saddle River, NJ, 1991).
[4] N. Lasry, J. Watkins, and E. Mazur, "Peer instruction: From Harvard to the two-year college," Am. J. Phys. **76** (11), 1066–1069 (2008).
[5] D. Meltzer and K. Manivannan, "Transforming the lecture-hall environment: The fully interactive physics lecture," Am. J. Phys. **70** (6), 639–654 (2002).
[6] L. Van Dijk, G. Van Der Berg, and H. Van Keulen, "Interactive lectures in engineering education," Eur. J. Eng. Educ. **26**, 15–28 (2001).
[7] S. Rao and S. DiCarlo, "Peer instruction improves performance on quizzes," Advanced Physiology Educ. **24**, 51–55 (2000).

[8] C. Turpen and N. Finkelstein, "Not all interactive engagement is the same: Variations in physics professors' implementation of Peer Instruction," Phys. Rev. ST Phys. Educ. Res. **5** (2), 020101-1–18 (2009).
[9] D. Duncan, "Clickers: A new teaching aid with exceptional promise," Astron. Educ. Rev. **5** (1), 70–88 (2006).
[10] I. Beatty, W. Gerace, W. Loenard, and R. Dufresne, "Designing effective questions for classroom response system teaching," Am. J. Phys. **74** (1), 31–39 (2006).
[11] S. Willoughby and E. Gustafson, "Technology talks: Clickers and grading incentive in the large lecture hall," Am. J. Phys. **77** (2), 180–183 (2009).
[12] M. James, F. Barbieri, and P. Garcia, "What are they talking about? Lessons learned from a study of peer instruction," Astron. Educ. Rev. **7** (1), 37–43 (2008).
[13] M. James, "The effect of grading incentive on student discourse in Peer Instruction," Am. J. Phys. **74** (8), 689–691 (2006).
[14] A. Strauss, *Constant Comparison: Qualitative Analysis for Social Scientists* (Cambridge U. P., New York, 1987).
[15] The clicker system used by Instructors 2 and 3 was manufactured by e-Instruction. Instructor 1 used the i-Clicker system.
[16] The F test is a test for the statistical significance of an observed difference between the means of two samples. The "*p*-value" is the probability that a statistical finding occurred by chance.
[17] A. Shamir, M. Zion, and O. Spector, "Peer tutoring, metacognitive processes and multimedia problem-based learning: The effect of mediation training on critical thinking," J. Sci. Educ. Technol. **17**, 384–398 (2008).
[18] N. Grover, "How to create successful discussion in science classrooms," Biochem. Mol. Biol. Educ. **35** (6), 397–403 (2007).
[19] J. Barell, *Teaching for Thoughtfulness: Classroom Strategies To Enhance Intellectual Development*, 2nd ed. (Longman, White Plains, 1995).
[20] L. Ding, N. Reay, A. Lee, and L. Bao, "Are we asking the right questions? Validating clicker question sequences by student interviews," Am. J. Phys. **77** (7), 643–650 (2009).
[21] J. Clement, "Expert novice similarities and instruction using analogies," Int. J. Sci. Educ. **20** (10), 1271–1286 (1998).
[22] D. Treagust and C. Smith, "Secondary students' understanding of gravity and the motion of planets," Sch. Sci. Math. **89** (5), 380–391 (1989).
[23] E. Bardar, E. Prather, K. Brecher, and T. Slater, "Development and validation of the light and spectroscopy concept inventory," Astron. Educ. Rev. **5** (2), 103–113 (2006).



Spectrometer. In the fall of 1957 I took a course in Optics and Wave Phenomena that was part of the physics major program at Amherst College. One of the experiments was on the operation and uses of a precision spectrometer made about 1900 by the Societe Genevoise in Switzerland. Young Tom Greenslade, on the right, and his laboratory partner were required to spend an hour becoming familiar with the instrument before starting the relatively simple experiment. The eleven inch divided circle can be resolved into one minute increments using the measuring microscopes. Fifty two years later I found that the spectrometer is still available for use in the laboratory. (Notes and photograph by Thomas B. Greenslade, Jr., Kenyon College.)