A national study assessing the teaching and learning of introductory astronomy. Part I. The effect of interactive instruction

Edward E. Prather
Department of Astronomy and Center for Astronomy Education, University of Arizona, Tucson, Arizona 85721

Alexander L. Rudolph
Department of Physics, California State Polytechnic University, Pomona, California 91768

Gina Brissenden and Wayne M. Schlingman
Department of Astronomy and Center for Astronomy Education, University of Arizona, Tucson, Arizona 85721

(Received 25 July 2008; accepted 15 December 2008)

We present the results of a national study on the teaching and learning of astronomy as taught in general education, non-science-major, introductory astronomy courses. Nearly 4000 students enrolled in 69 sections of courses taught by 36 different instructors at 31 institutions completed (pre- and post-instruction) the Light and Spectroscopy Concept Inventory (LSCI) from Fall 2006 to Fall 2007. The classes varied in size and were from all types of institutions, including 2- and 4-year colleges and universities. Normalized gain scores for each class were calculated. Pre-instruction LSCI scores were clustered around ~25%, independent of class size and institution type, and normalized gain scores varied from about −0.07 to 0.50. To estimate the fraction of classroom time spent on learner-centered, active-engagement instruction we developed and administered an Interactivity Assessment Instrument (IAI). Our results suggest that the differences in gains were due to instruction in the classroom, not the type of class or institution. We also found that higher interactivity classes had the highest gains, confirming that interactive learning strategies are capable of increasing student conceptual understanding. However, the wide range of gain scores seen for both lower and higher interactivity classes suggests that the use of interactive learning strategies is not sufficient by itself to achieve high student gain.
[DOI: 10.1119/1.3065023]

I. INTRODUCTION

The *American Journal of Physics* has published many articles on the work done in the field of physics education research (PER). However, very little has been published on work done in the field of astronomy education research (AER). Though the types of questions the researcher asks, and how the research is conducted, are very similar in both fields, the central population of students and courses being researched are very different.[1–5] In PER, the student-focus has been on science majors taking calculus-based physics (and to a lesser degree algebra-based physics). In AER, the student-focus has been on students (primarily non-science majors) taking a general education, college level, introductory astronomy course, which we will refer to as Astro 101. Though these two populations of students are different, previous research[2,3,6–17] does highlight similarities in instructional difficulties that affect both populations.

Some of the difficulties include that both physics and astronomy students do not achieve deep conceptual understanding of physics or astronomy topics through traditional, lecture-based instruction alone,[2,3,6–13] and that they leave their physics and astronomy courses with little to no improvement in their attitudes toward, values about, or interests in science.[14–17] This research also demonstrates that student understanding is significantly increased when learner-centered teaching strategies are used in conjunction with a metacognitive instructional approach which allows students to assess their own understanding.[18] These results are consistent with the findings of decades of research on how humans learn beyond the realm of teaching and learning in physics and astronomy.[19]

Differences in these populations begin at a very basic level—science vs. non-science majors—and extend beyond this to the student's role in society when outside of our Astro 101 courses and college. Annually, nearly 250 000 students take Astro 101, with approximately 100 000 of these students taking their courses at 2- and 4-year institutions that do not offer degrees in astronomy or physics.[20] Again, these are mostly non-science majors[32]—they are our society's future lawyers, journalists, business people, and politicians. Their motivations for taking Astro 101, their motivations to learn the content of the course, as well as their preexisting math and science knowledge and skills are different than those of students who are science majors taking algebra- and calculus-based physics. Perhaps most importantly, these are not our future scientists, but many are our future *teachers*–40% of students taking introductory science courses state that they intend to become licensed teachers.[21]

In the PER community, dialogue about the nature of teaching and learning in physics, and the role of the physics instructor in creating an effective learning environment, took a major step forward driven by the results from the study conducted by Hake[8] (we will reference this study throughout the rest of this paper as "Hake"). The evidence presented by Hake clearly showed that many students were not achieving in physics at the level one would expect, and the primary determining factor affecting gains was whether or not interactive learning strategies were used. These results led to a sort of revolution in physics instruction. Physics instructors with no experience in PER were having conversations regarding pedagogy that were driven by an emphasis on research results on student understanding. In addition, the physics community started treating their teaching more scientifically, using the results of the Hake study to inform their instruction. Another difference between the PER and AER communities is that, thus far, no similar study has been conducted in astronomy to provide the same impetus for a revolution.

Hake used the Force Concept Inventory to assess the success of physics instruction on a national scale. Until recently, the AER community has had no similar instrument or concept inventory—namely one that covered a central topic of the Astro 101 curriculum (like Newton's laws are to first semester physics courses) and that had been shown to be valid and reliable at differentiating student understanding and instructional type. Over the past few years, concept inventories have been developed on lunar phases,[22] properties and formation of stars,[6] the greenhouse gas effect,[23] and light and spectroscopy.[7] Hake also developed a survey instrument to document whether a class should be classified as a traditional lecture course or was one that used interactive learning strategies. Again, the AER community has had no such instrument. To address the lack of such an instrument, we developed the Interactivity Assessment Instrument (IAI), designed to quantify the amount of interactive engagement the students received (beyond lecture) in the classrooms included in our study.

In an attempt to motivate a similar revolution in AER, we conducted a national research project to study how the use of interactive learning strategies affects learning in Astro 101 classes. For our study, we chose to use the Light and Spectroscopy Concept Inventory (LSCI) as our instrument because of how central these topics are to all astronomy—light is the fundamental carrier of astronomical information and spectral features serve as the "fingerprints" of the universe.[24] In addition, research has shown that Astro 101 instructors consider the nature of light and the electromagnetic spectrum as the most important, and the most taught, topic in Astro 101.[25,26] In addition to the 26 questions of the LSCI, we asked each student to answer 15 demographic questions.

In this paper we discuss the results of our LSCI study—gain score distributions, similarities and differences related to Hake's study, and what this tells us about our students and the nature of instruction in Astro 101 classrooms. We also discuss the development of, and results from, our use of the IAI, including the extent to which gain scores on the LSCI can be correlated with learner-centered instruction. Subsequent papers will look at the demographic data we collected along with the LSCI, including how student conceptual gains are related to various demographic categories.

Section II outlines the study methodology, including details of the target population, the LSCI instrument, the method of data collection and analysis, and the development and use of the IAI. Section III outlines the results of our study, including a description of the dataset, and our analysis of the relation between gain scores and pre-instruction scores, class size, institution type, and level of interactivity in the classroom, as measured by the IAI. Section IV discusses the main conclusions of the study.

II. STUDY METHODOLOGY

A. The target population

For this study, our goal was to gather data from the Astro 101 community that represented the widest possible range of instructional techniques and classroom settings, the latter including class size, institution type, and geographical distribution. The population in this study was the students enrolled in a general education, non-science-major, introductory astronomy course (commonly referred to as "Astro 101") and their instructors. The participating instructors were recruited through two different invitations. One was a posting made to the Center for Astronomy Education (CAE) academic listserv for astronomy teaching and learning.[36] In addition, emails were sent to all members of the CAE mailing list, which includes past participants of the CAE professional development workshops. These workshops focus on developing instructors' pedagogical content knowledge[27–30] in an effort to improve their implementation of learner-centered teaching strategies. Additional details about the number of students, as well as the number and types of institutions, are presented in the dataset portion of Sec. III.

B. The instrument: The Light and Spectroscopy Concept Inventory

For this study, we chose as our instrument the Light and Spectroscopy Concept Inventory (LSCI), a 26-question multiple-choice diagnostic test developed by Bardar *et al.*[7] The LSCI, which is research proven, valid, and reliable, was designed to measure the change in a student's conceptual understanding of light and spectroscopy, topics fundamental to the learning that will occur throughout an Astro 101 course.[7,24,31] It is our assertion that using an instrument focused on central topics to the entire course allows one to evaluate the overall effectiveness of instruction in that classroom.

Concepts addressed in the LSCI include[7]

- the nature of the electromagnetic spectrum, including the interrelationships of wavelength, frequency, energy, and speed;
- interpretation of Doppler shift as an indication of motion rather than color of an object;
- the correlation between peak wavelength and temperature of a blackbody radiator;
- relationships between luminosity, temperature, and surface area of a blackbody radiator; and
- the connection between spectral features and underlying physical processes

C. Data collection and reduction

Once recruited, instructors were sent a copy of the LSCI (including the demographic questions), plus sufficient Scantron™ forms to administer the LSCI. The instructors were directed to ask students to take the LSCI voluntarily before instruction began (pre-test) and again after all instruction on the concepts of light and spectroscopy had occurred, at the end of the course (post-test). The Scantron™ forms were then returned to the authors for analysis. The data were collected over a 17-month period from Fall 2006 through Fall 2007.

The Scantron™ forms were first visually inspected. For a Scantron™ form to become part of the dataset, it had to have

Fig. 1. Geographical distribution of the 30 participating institutions in the United States.

no more than 4 of the 26 LSCI questions left blank and could not have any obvious "geometric" pattern of answers (e.g., all A's, a zigzag pattern), indicating the student was not trying earnestly to answer the questions to the best of their ability. These rejections formed a very small percentage (less than 2%) of all forms collected. With the remaining data set, the 26 LSCI questions were scored and a pre-instruction score (pre-test) and post-instruction score (post-test) were recorded for each student. From the average pre- and post-test percentages of each class, a normalized gain, $\langle g \rangle$, was calculated:

$$\langle g \rangle = \frac{\langle \text{post } \% \rangle - \langle \text{pre } \% \rangle}{100 - \langle \text{pre } \% \rangle}.$$

Normalized gain is the ratio of the percentage gain achieved to the possible improvement that could be achieved, as determined by the pre-test score.

D. Interactivity levels in the classroom

The Interactivity Assessment Instrument (IAI) was developed to help provide a first-order indicator of the connection between the gain in students' conceptual understanding and the type of instruction they received, in particular the extent that traditional lecture and different interactive learning strategies were used in the classroom. This eight-item instrument was explicitly designed to enable us to quantify the amount of time spent on interactive learning strategies that occurred in each classroom but does not directly provide further insight into the effectiveness of an instructor's implementation of particular interactive learning strategies. We use the phrase "interactive learning strategies" to identify those strategies that have been designed to intellectually engage students in critical thinking (and increase their conceptual understanding) while working in a collaborative learning group with one or more peers. The strategies identified in the IAI (Think-Pair-Share questions, Lecture-Tutorials, and Ranking Tasks) have undergone research-validated studies within the astronomy and physics education communities to show that they are capable of significantly increasing students' conceptual understanding.[9,10,33] In addition, interactive learning strategies identified in the IAI are appropriate for use in all Astro 101 classes, small and large. The IAI is included as the Appendix to this article.

From a design and research perspective, the first four questions (e.g., "How many total contact hours do you have each week with your class?") were created to determine the total number of possible hours of instruction available in each class over the course of the term (semester or quarter). The fifth and sixth questions were designed to elicit how often different interactive learning strategies were implemented into each class or throughout the instructional term. The seventh question was designed to determine how often students were asked a question or asked to make a prediction

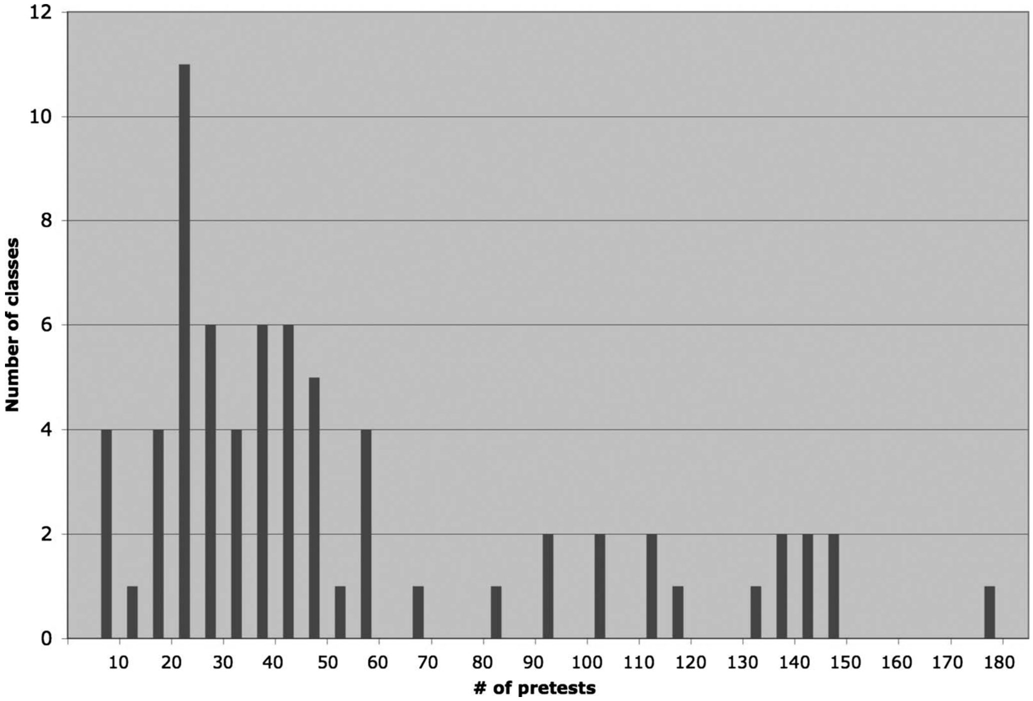

Fig. 2. Histogram of the number of classes binned by the number of pre-tests.

Table I. Summary of study data sorted by institution type.

Institution type	Institutions[a]	Instructors	Sections	Pre-tests	Post-tests	Post-tests (matched)
All	**31**	**36**	**69**	**3729**	**2577**	**1970**
2–year Colleges	12	14	34	1045	672	491
4-year Bachelor Colleges	3	3	4	108	107	88
4-year Bachelor/Masters Universities	10	12	19	1267	900	737
Research Universities	6	7	12	1309	898	654

[a] These 31 institutions include the 30 plotted in Fig. 1 plus one in Ireland.

by themselves, without collaboration with other students. The final question of the IAI asked instructors to provide a description of any instructional strategies that were not directly asked about in the prior questions of the IAI, and to provide the frequency they were used and the amount of in-class time spent on each of these strategies. Unfortunately the final question did not provide the additional insights into classroom instruction that we had hoped it would. None of the instructors in the study provided responses with sufficient detail for us to determine how often one of these instructor-provided strategies was used, or how much time should be allocated out of the term's instruction for its use.

The information provided by instructors on their IAI was then used to calculate an Interactivity Assessment Score (IAS) in the following way: a nominal time was assigned to each instructional strategy, based on our knowledge of the typical times spent on such activities. The times assigned for each activity were Lecture Tutorial and Ranking Task (15 min for either), Think-Pair-Share Question (3 min), and Question or Prediction students made alone (1 min). The amount of time spent during the term on interactive instruction was then added up based on each instructor's responses and then divided by the total available time for instruction in the term. Hence, the IAS provides a rough estimate of what fraction of the total available in-class instructional time was spent using interactive instruction.

Instructor IAS's ranged from 0% to 49%, suggesting that our instrument was successful at distinguishing different levels of interactivity in the classroom. However, the IAS is only a first-order measurement of the time spent on interactive instruction and, by itself, provides no detail as to the quality of the implementation or engagement in these classrooms.

In an effort to be sure that instructors would not alter the instruction in their courses in response to the IAI, and to ensure that their answers were closely matched to the actual instruction that occurred in their courses, instructors completed their surveys near or shortly after the end of instruction for the term. We did not tell the instructors how we were using these data, and we were careful to write questions that did not reveal what we wanted to infer from the data. We were concerned that if instructors knew the purpose of the questions they might bias their results, e.g., providing responses suggesting their classrooms were more interactive than they were in reality.

III. RESULTS

A. The data set

A total of 36 instructors teaching 69 sections at 30 different institutions across the U.S. (plus one in Ireland) participated in the study. These institutions span all four types of colleges with designations of Associate (2-year) Colleges, Baccalaureate Colleges (4-year primarily Bachelor granting Colleges), Master's Colleges and Universities (4-year primarily Masters and Bachelors granting Universities), and Doctorate-granting Universities (Research Universities). The geographic distribution of participants is national in coverage and is shown in Fig. 1

Figure 2 shows a histogram of the number of classes binned by the number of pre-test responders (which we use as a proxy for class size). The class sizes ranged from 6 pre-test responders up to 180 pre-test responders, with a mean of 54 and a median of 38. We have a total of 3729 pre-test responders, 2577 post-test responders, and 1970 responders for which we were able to match a pre-test to a post-test. Table I provides a summary of the total dataset collected sorted by institution type. Table II provides the same information sorted by class size.

As a first assessment of the dataset as a whole, we calculated the pre- and post-test scores and percentages for each class section for all students who took the LSCI, regardless of whether we were able to match their pre- and post-tests. Thus, we included students who started the class but did not finish for some reason as well as including some students who were present on either the day that the pre- or post-test was given, but not both. However, we will argue shortly that the scores of all students who took the pre-test are statistically indistinguishable from the pre-test scores of those who took both the pre- and the post-test (i.e., were matched) and therefore stayed for the entire course.

B. Gain versus pre-test scores

From the average pre- and post-test percentages, we calculated a normalized gain for each class section. In Figs. 3 and 4, we plot this normalized gain against the average pre-test percentage (again for all students who took the pre-test), for each of the 69 class sections in the study. Figure 3 pre-

Table II. Summary of study data sorted by class size (N=number of pre-tests).

Class size	Sections	Pre-tests	Post-tests	Post-tests (matched)
All	**69**	**3729**	**2577**	**1970**
Very small ($N<25$)	15	252	176	132
Small ($N=25-49$)	31	1092	817	598
Medium ($N=50-99$)	10	674	457	380
Large ($N\geq 100$)	13	1711	1127	860

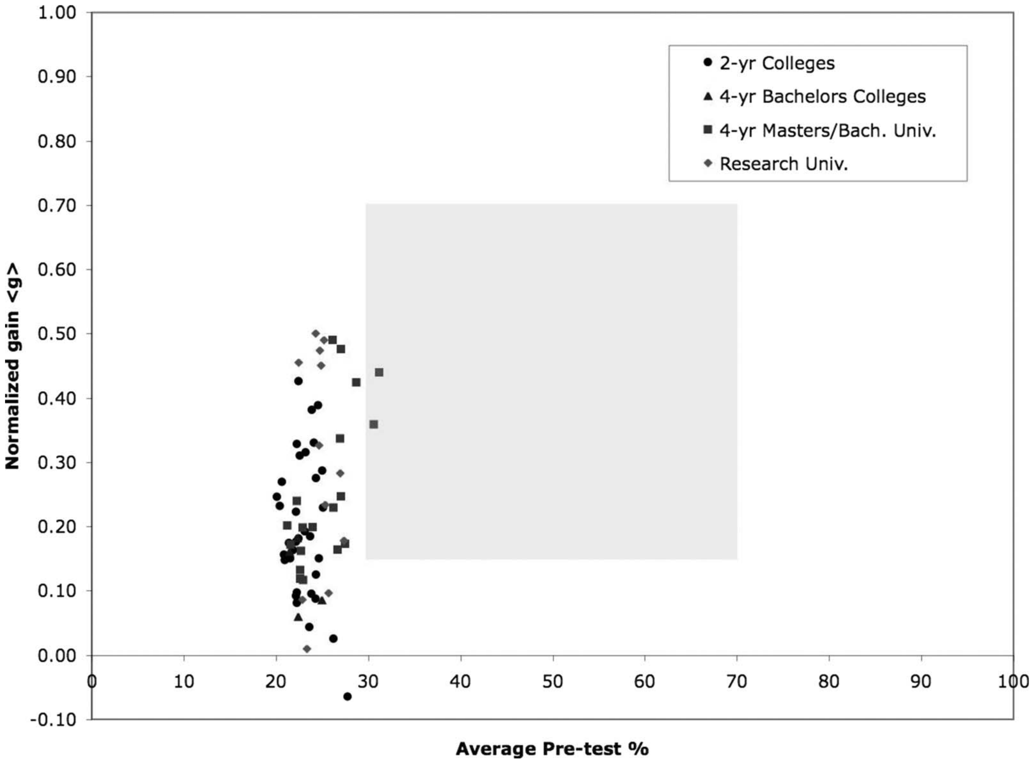

Fig. 3. Plot of normalized gain $\langle g\rangle$ vs. average pre-test percentage for all 69 class sections in the study. The different symbols represent the four different types of institutions at which the classes were taught. The shaded region represents the range of pre-test scores and normalized gain seen by Hake for college-level classes in his study of introductory physics classes.

sents this data identified by the type of institution at which the class was taught; in contrast, Fig. 4 presents the data identified by the size of the class as measured by the number of pre-test scores. Figures 3 and 4 are different from Fig. 1 in Hake's study. One critical difference is that we plot normalized gain, $\langle g\rangle$, on the vertical axis whereas Hake plotted %⟨Gain⟩ (which is the difference between %post-test and %pre-test) on his vertical axis. In both our study and Hake's study the variable which is most representative of student learning is normalized gain $\langle g\rangle$. Due to Hake's choice to plot %⟨Gain⟩ on the vertical axis, constant normalized gain must be represented as lines of negative slope on his graphs. We chose, for clarity, to explicitly plot normalized gain $\langle g\rangle$ on our vertical axis. In this case constant normalized gain $\langle g\rangle$ would be represented as a horizontal line. To aid in the comparison of the distribution of scores between our study and Hake's, in Figs. 3 and 4 we have provided a shaded region to represent the range of pre-test and normalized gain scores $\langle g\rangle$ for the college-level classes in Hake's data.

A number of important results are evident from these two plots:

(1) The range of pre-test percentages is surprisingly narrow, clustered around 25% ($24\pm2\%$), regardless of institution type or class size. This is very different from Hake's study, in which pre-test percentages ranged from a low of 30% to as high as 70% (as indicated by the width of the shaded box in Figs. 3 and 4). This suggests that there is a fundamental difference in the student population taking Astro 101 from those taking introductory college-level physics classes. We can infer that many of the physics students in Hake's study had significant instruction on the topics assessed by the FCI, or at least prior knowledge of these topics, before taking their first college-level physics course. By contrast, the students in Astro 101 courses, who are drawn from all majors (but are overwhelmingly non-science majors),[32] seem to have not had significant exposure to, nor prior knowledge of, the topics addressed by the LSCI prior to taking Astro 101, or they have had instruction that had little effect on their understanding. It is worth noting that, because most questions on the LSCI have four to five possible answers, a pre-test score of 25% is roughly consistent with guessing.

(2) The highest gain scores (~0.50) are somewhat lower than found in the Hake study using the FCI (for which the highest gain scores were ~0.70), suggesting that the LSCI is nontrivial and is truly testing students' understanding of difficult-to-grasp concepts. This result, combined with the fact that the gain scores in our study show a large spread, from about (−)0.07 to 0.50, suggests that the LSCI is a valid instrument capable of measuring the effectiveness of teaching the topic of light in Astro 101 classes.

(3) Gain scores do not depend on either *institution type or class size*. That is, no students from *any* institution type appear to perform at a higher level than any other institution type, as measured by gain score. Further, there appears to be no advantage to being a student in a large or small class. This latter point might seem surprising as one might predict that teaching a very small class ($N<25$ students) might allow for a level of personalized instruction, which would give an advantage to students in those classes, but this was clearly not the case as

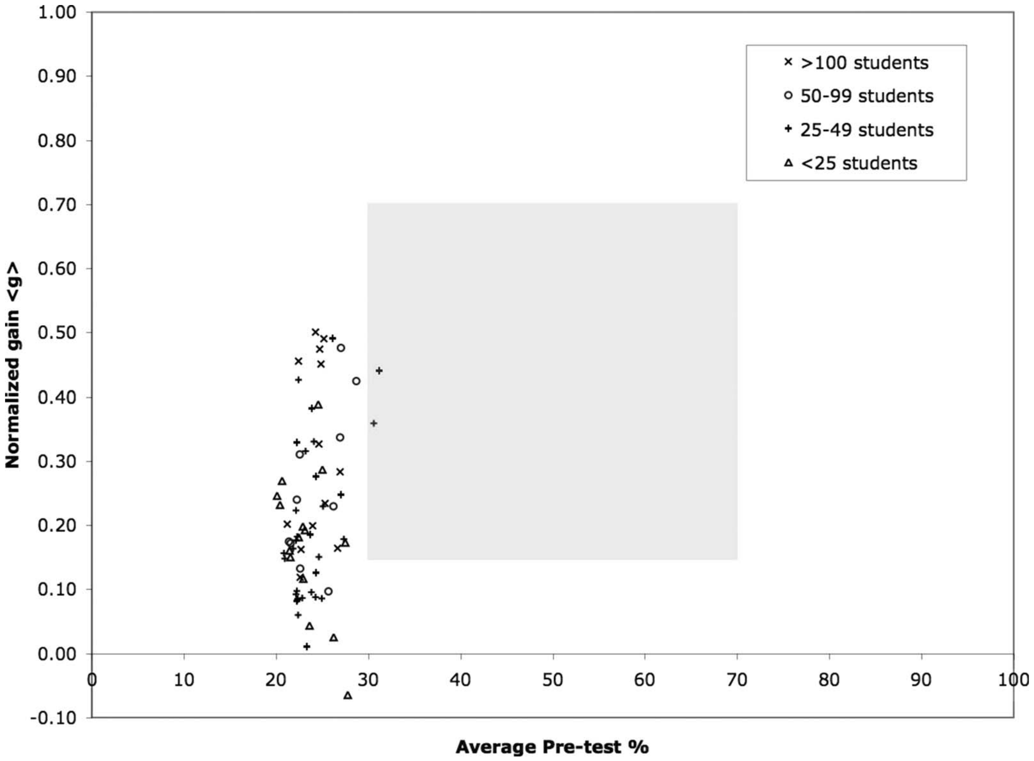

Fig. 4. Plot of normalized gain $\langle g \rangle$ vs. average pre-test percentage for all 69 class sections in the study. The different symbols represent different class sizes (based on number of pre-tests). The shaded region represents the range of pre-test scores and normalized gain seen by Hake for college-level classes in his study of introductory physics classes.

documented by this study. We will explore in more depth the dependence of gain on the parameters of class size and institution type, as well as on the relationship between gain and interactivity in the following sections.

For the remainder of our analysis of the LSCI data we chose to limit our investigation to classes with at least 25 pre-test scores ($N \geq 25$). We do this for three reasons. First, it is our assertion that the teaching and learning that occur in classes with a very small number of students can be quite different (bordering on personalized instruction in some cases) from what is possible to accomplish in large enrollment classes. Very small classes allow for the use of interactive learning strategies that are not viable in larger classes, and which the IAI was not designed to measure. The IAI was purposely designed to measure those interactive learning strategies that are appropriate to all class sizes, large and small. However, it was not designed to measure interactive learning strategies that are only possible in very small classes, but not large classes. Second, very small classes can provide less reliable statistical results. Third, looking at Fig. 2, there is clearly a peak in the distribution of class sizes between 25 and 30 pre-test scores; hence, choosing $N \geq 25$ allows us to keep the largest number of classes in our data set and further supports our decision to use this threshold. Using $N \geq 25$ pre-test scores as the limiting factor removed only 15 class sections from our data. It is worth noting that these 15 sections contained a total of only 252 students; thus, we lost fewer than 7% of our total student population. This leaves 54 class sections within the dataset with at least 25 pre-test scores ($N \geq 25$).

For the remainder of this article, we only consider matched data for 52 of these 54 class sections of $N \geq 25$. Using a *t*-test comparing the unmatched students (those with only a pre-test *or* a post-test) and matched students (those with *both* a pre-test and a post-test), we were able to show that for these 52 sections, the pre-test scores of the two populations were equivalent at a 95% confidence level ($p > 0.05$). In addition, a *t*-test comparing the gain scores calculated from only the matched students to the gain scores of the entire data set (both matched and unmatched) showed that the two calculations of gain were statistically equivalent at a 95% confidence level ($p > 0.05$), for the same 52 class sections. Tables III and IV provide a summary of the data for the remaining 52 class sections ($N \geq 25$ pre-tests), sorted by institution type and class size, respectively.

In Figs. 5 and 6, we plot the normalized gain versus average pre-test percentage for the "matched" data from the 52 classes with $N \geq 25$ pre-tests. These new plots confirm the

Table III. Summary of study data for classes of $N \geq 25$ pre-tests sorted by institution type.

Institution type	Sections	Pre-tests	Post-tests	Post-tests (matched)
All	**52**	**3238**	**2226**	**1711**
2-year Colleges	23	862	550	398
4-year Bachelor Colleges	3	101	100	82
4-year Bachelor/Masters Universities	14	966	678	577
Research Universities	12	1309	898	654

Table IV. Summary of study data for classes of $N \geqslant 25$ pre-tests sorted by class size.

Class size	Sections	Pre-tests	Post-tests	Post-tests (matched)
All	**52**	**3238**	**2226**	**1711**
Small (N=25–49)	31	1092	817	598
Medium (N=50–99)	10	674	457	380
Large ($N \geqslant 100$)	11	1472	952	733

results described earlier regarding the entire dataset, inferred from Figs. 3 and 4, namely that the type of institution and the number of students in a course are not determining factors as to whether students will acquire an understanding of the nature of light as related to astronomy and measured by the LSCI. While some Astro 101 instructors are of the belief that achievement might be higher at 4-year colleges and universities versus 2-year community colleges, or that interactive learning strategies won't work at their particular type of college or with their students, our evidence suggests these beliefs are not consistent with the results from this study of Astro 101 courses around the country.

Since institution type and class size seem to have no correlation with student learning, and because we can assume that these 52 classes are not all taught identically, we can conclude that the different gains achieved are related to the effectiveness of the teaching and learning that students' experience in their classes. We will now investigate the degree to which the amount of interactive instruction used in a class is related to the level of understanding achieved, as measured by normalized gain.

C. Gain versus interactivity level

As we described in Sec. II D, the Interactivity Assessment Instrument (IAI) was used to provide a measure of how much in-class time was spent on instruction designed to engage students in their learning beyond what is accomplished from lecture alone. We calculated Interactivity Assessment Scores (IAS's) ranging from 0% (classes that spend 100% of their time in lecture) to 49% (classes that spend approximately half of their time engaged with interactive learning strategies). While the instructor-supplied responses to the IAI were helpful at quantifying the percent of class time dedicated to interactive learning strategies, we would not claim that the IAS is a measure of the quality of implementation of these interactive learning strategies. Furthermore, given its current design, the IAI does not provide enough detail to allow comparison of the amount (or percent) of class time spent on each interactive learning strategy used by instructors over a term. Therefore, we are unable to use the IAS of a single class to compare how different learning strategies affect the gain in understanding achieved by the students in that class. However, if we plot normalized gain vs. IAS, as shown in Fig. 7, it is possible to infer the extent to which increased class time spent on interactive learning strategies is related to students' understanding of LSCI topics.

In looking at Fig. 7, it is important to notice there is a distinct difference in the distribution of normalized gain

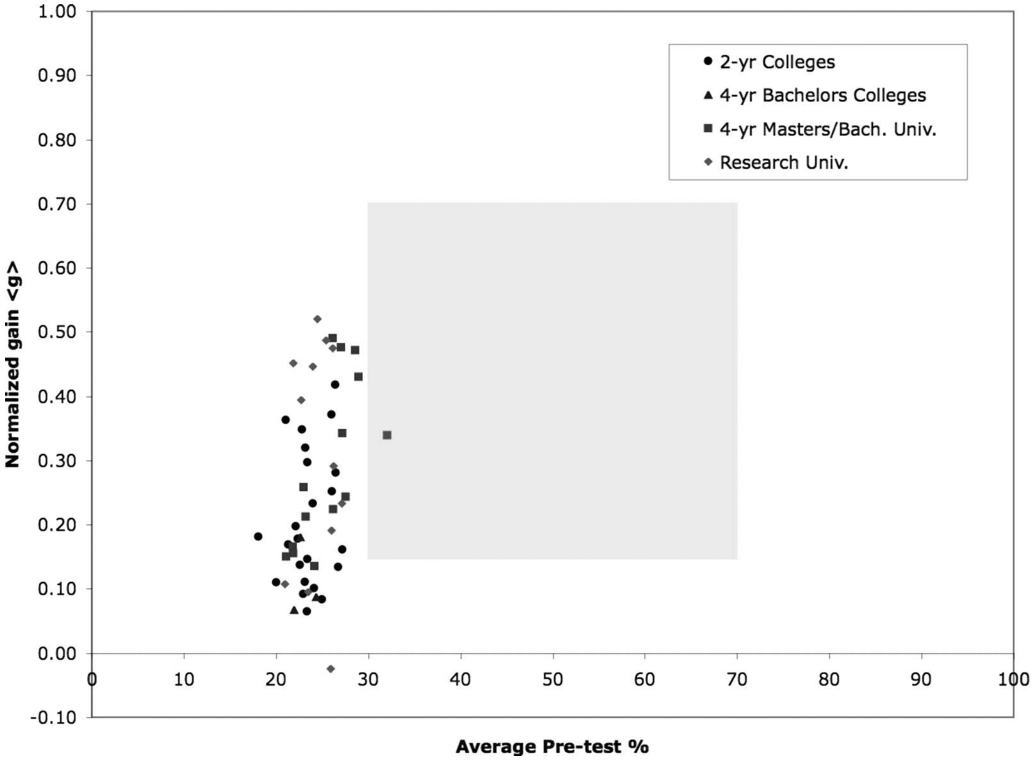

Fig. 5. Plot of normalized gain $\langle g \rangle$ vs. average pre-test percentage for the 52 class sections with $N \geqslant 25$ pre-tests. Pre-test scores and normalized gains are only for students who took both the pre-test and the post-test ("matched" student data). The different symbols represent the four different types of institutions at which the classes were taught. The shaded region represents the range of pre-test scores and normalized gain seen by Hake for college-level classes in his study of introductory physics classes.

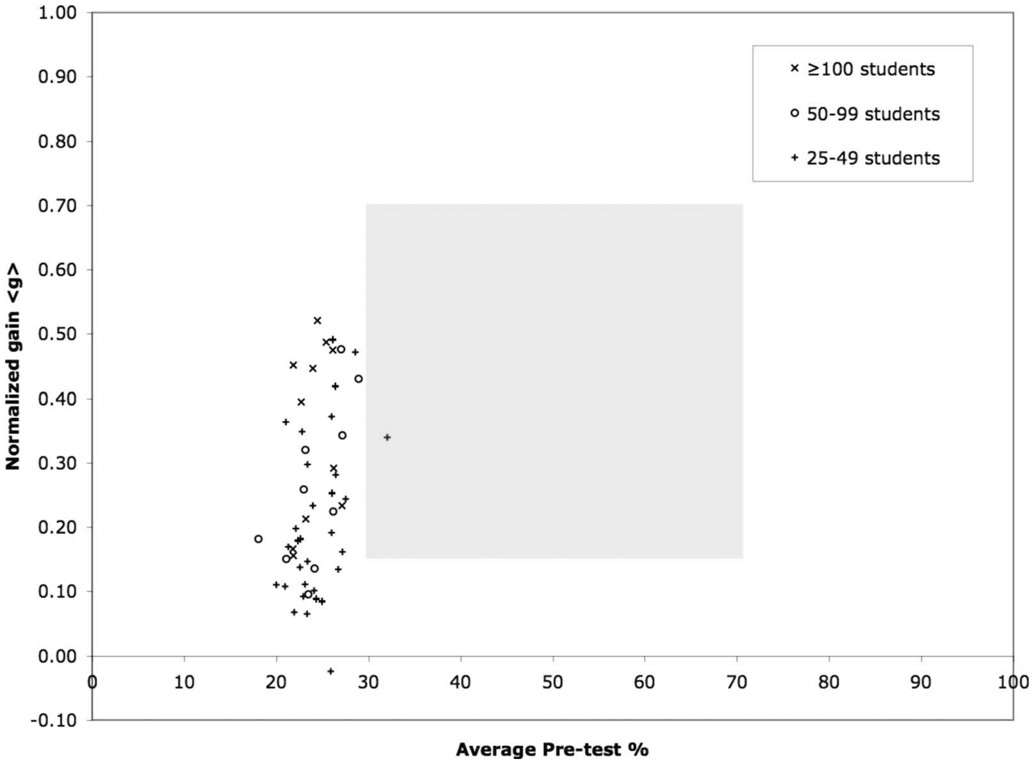

Fig. 6. Plot of normalized gain ⟨g⟩ vs. average pre-test percentage for the 52 class sections with $N \geqslant 25$ pre-tests. Pre-test scores and normalized gains are only for students who took both the pre-test and the post-test (“matched” student data). The different symbols represent different class sizes (based on number of unmatched pre-tests). The shaded region represents the range of pre-test scores and normalized gain seen by Hake for college-level classes in his study of introductory physics classes.

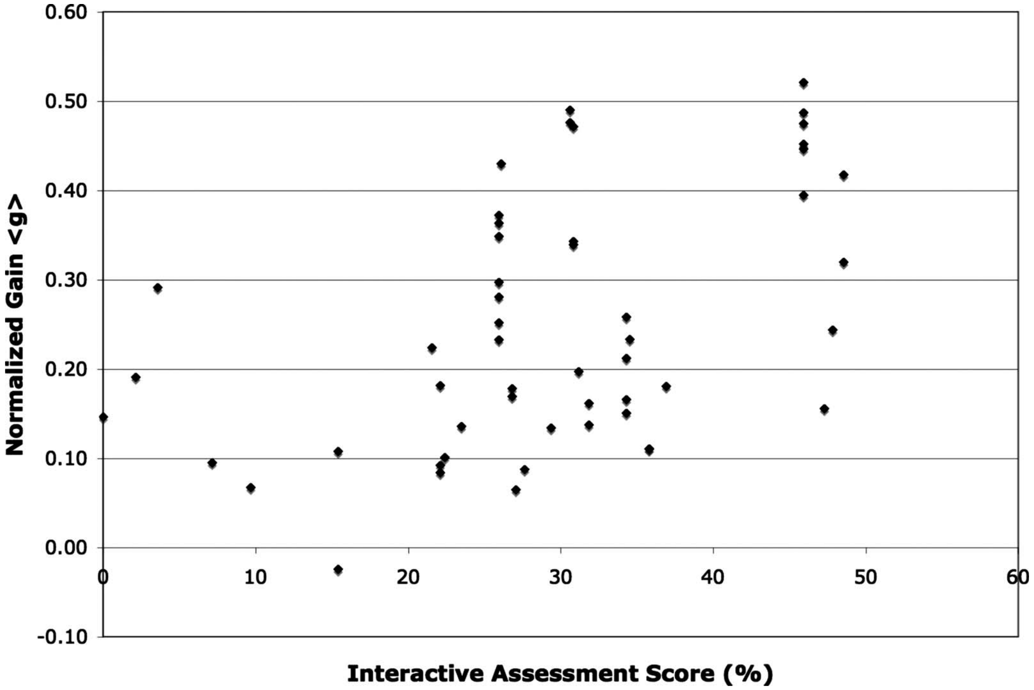

Fig. 7. A plot of normalized gain ⟨g⟩ vs. Interactive Assessment Score (IAS), derived from instructor responses to the IAI, for the 52 class sections with $N \geqslant 25$ pre-tests.

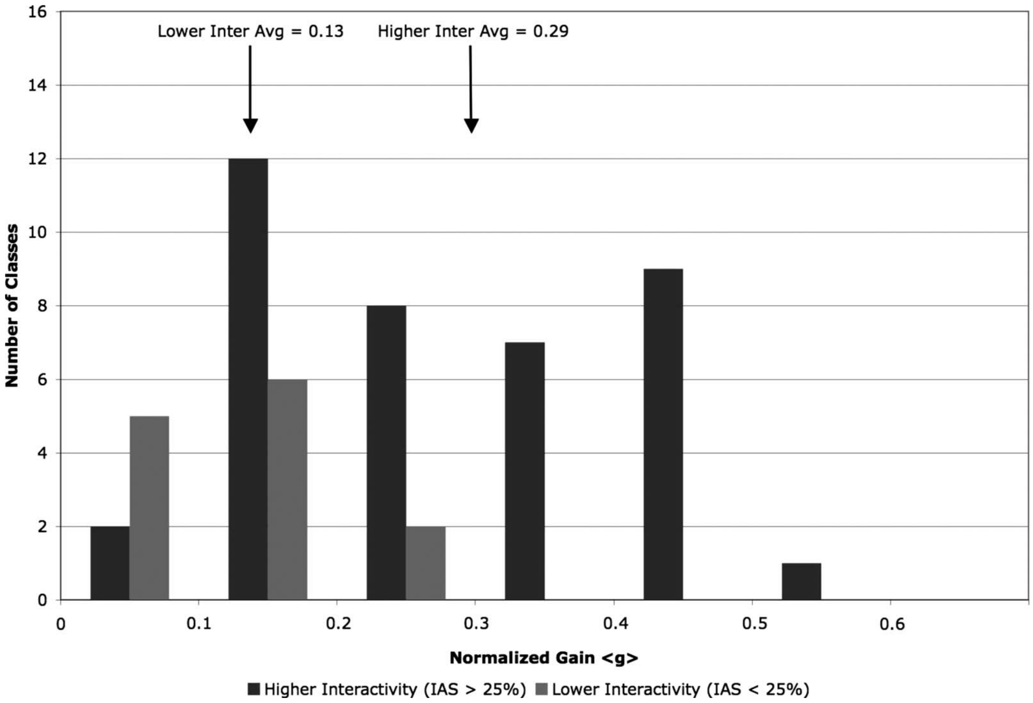

Fig. 8. Histograms of the number of classes binned by normalized gain for both lower interactivity classes $(\text{IAS} < 25\%)$ and higher interactivity classes $\text{IAS} > 25\%$. The average value of the normalized gain for the low and high interactivity class distributions is indicated with an arrow at the top of the plot. The lower interactivity average is 0.13, and the higher interactivity average is 0.29.

scores between classes with an IAS below 25% (lower IAS) and those between 25% and 50% (higher IAS). Of the 13 lower IAS classes (taught by ten instructors), note that none achieve a normalized gain above 0.30, whereas the 39 higher IAS classes have gains ranging from 0.05 to 0.52, with 22 classes (56%) below 0.30 and 17 classes (44%) above 0.30. Of the 19 instructors who taught the 39 higher IAS classes, 6 instructors had classes with gains above 0.30, 12 had classes with gains below 0.30, and 1 instructor had classes both above and below a gain of 0.30. Coupled together, the above results lead to the following inferences:

(1) Since only higher IAS classes achieved higher gains ($\langle g \rangle$ greater than 0.30), we infer that interactive learning strategies are capable of improving students' conceptual understanding of the concepts of light and spectroscopy in Astro 101. The break in gain scores at an IAS of 25% indicates that classes which don't dedicate at least one fourth of their class time to interactive learning strategies may never achieve gains on the LSCI of greater than 0.30.

(2) Although the use of interactive learning strategies appears to have the ability to help students improve their understanding of LSCI topics, simply spending more time on interactive learning strategies did not ensure that higher IAS classes would achieve gains of 0.30 or more. This is evidenced by the significant number of classes (22) with an IAS above 25% but a $\langle g \rangle$ below 0.30 (approximately 50% of the classes with IAS above 25%). While we cannot be certain of the cause of the spread in gain scores seen in the higher interactivity groups, we believe that the quality of an instructors' implementation of the interactive learning strategies used in a class is very likely related to this observed spread in gain scores. Note that other studies have found support for the assertion that quality of implementation of instructional strategies can have a significant influence on students' learning gains.[34,35]

Although our results support the idea that interactive learning strategies do have a positive impact on students' conceptual understanding of light and spectroscopy, the small number of lower IAS classes leaves open the possibility that a broader range of gain scores might be achieved in lower IAS classes (including lecture-only classes) but our sample simply does not include any "high-achieving" lower-IAS classes. To further investigate the relationship between normalized gain and level of interactivity, we plotted histograms of the lower and higher IAS groups, treated as two distinct distributions, shown in Fig. 8. The value of the average (mean) normalized gain for each distribution is indicated with an arrow at the top of the plot.

The lower IAS distribution seems to peak at a gain between 0.10 and 0.20, with an average gain for this group of 0.13, while the higher IAS distribution is more spread out, with gains from 0 to 0.50, but has a much higher average gain of 0.29. To test if these distributions are statistically different, we performed an independent sample t-test and found that they were ($p < 0.001$). In addition, these two distributions are seen as being distinctly different as evidenced by their very large (Cohen's d) effect size of 1.41. These results suggest that the differences seen between the two groups are not random, but rather are due to the real ability of higher IAS classes to achieve higher gain scores. Nonetheless, we advocate collecting more data from lower IAS classes to further test this hypothesis.

IV. CONCLUSIONS

We have conducted a national study on the teaching and learning of astronomy in the general education, non-science-major, introductory astronomy course (Astro 101). We tested students' conceptual gains in understanding light and spectroscopy, two of the central and fundamental topics of astronomy, by administering the Light and Spectroscopy Concept Inventory (LSCI), both pre- and post-instruction. We also developed the Interactivity Assessment Instrument or IAI, which was given to the instructors participating in our study, to gauge what fraction of class-time was spent using interactive instruction.

The data from this study suggest that the use of interactive learning strategies in Astro 101 classrooms can have a large effect on the conceptual gains students achieve. From the data we also infer that simply dedicating a larger percentage of class times to interactive learning strategies does not neccassarily translate into higher gain scores, as evidenced by the wide spread in gains seen in both lower and higher interactivity classrooms. In addition, since differences in gain between the various classes in our study were not correlated with such factors as institution type or class size, we suggest that it is the proper implementation of interactive learning strategies that is key to achieving higher gains in student learning.

ACKNOWLEDGMENTS

The authors would like to thank Michael Greene and the Navigator Public Engagement Program (NASA JPL) and Michelle Thaller and the Spitzer Education and Public Outreach Program (CalTech) for their generous and continued funding, as well as their vision, in helping make CAE possible. The authors would also like to thank the NSF and Duncan McBride for their very generous funding which has allowed CAE through the CCLI Phase III Collaboration of Teaching Scholars (CATS) program to carry out this research. In addition, the authors would like to thank the many undergraduate research assistants who mailed out hundreds of envelopes, ran thousands of Scantrons™, and assisted with initial data reduction: Tori Carr, Tatsunobo Yoshimoto, Ran Kumada, Chiho Kitajima, Motohura Kawazu, and Edward Montiel. Finally, the authors would like to thank all the members of the CAE community who cared enough about their students' learning to participate in this study, and the thousands of Astro 101 students who freely gave of their time to participate. Your conscientious efforts, the authors believe, will help them all be better Astro 101 instructors. The authors could never have done this without you. This material is based upon work supported by the National Science Foundation under Grant No. 0715517, a CCLI Phase III Grant for the Collaboration of Astronomy Teaching Scholars (CATS). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

APPENDIX: THE INTERACTIVITY ASSESSMENT INSTRUMENT

LSCI Instructor Survey Date:

Name: Term LSCI Administered:
Institution Name: Phone Number:

1. Is your institution on a **quarter** or **semester** system?
2. How many total contact hours do you have **each week** with your class? (Do not include office hours.)
3. How many hours do you spend each week in the following instructional setting?
    a. Lecture-style setting, which may include questions and activities
    b. Recitation or discussion section
    c. Doing traditional laboratory investigations and astronomical observations
    d. Other (please specify)
4. How many hours during the **term** are used for **exams** (not counting the final exam)?
5. During the **term**, how many times do students work in groups on *collaborative learning* activities, excluding traditional labs (e.g., Lecture-Tutorials, Ranking Tasks, Case Studies)? Provide a range if necessary.
6. During a **typical class**, how many times do you pose a question that requires your students to work together (e.g., Think-Pair-Share or Concept Test-type question)? Provide a range if necessary.
7. During a **typical class**, how many times are students asked to answer a question or make a prediction by themselves, without working together? Provide a range if necessary.
8. During a **typical class**, are there any other things that you do to promote student learning beyond lecture? For each (use the back-side of this sheet if you need more room);
    a. Provide a description
    b. State how many times **in a term** this occurs.

[1] A. Fraknoi and S. C. Wolff, “Astronomy Education Review: A Five-Year Progress Report and Thoughts about the Journal’s Future,” Astr. Ed. Rev. **5**(2), 1–4 (2007).

[2] J. M. Bailey and T. F. Slater, “Resource Letter AER-1: Astronomy Education Research,” Am. J. Phys. **73**(8), 677–685 (2005).

[3] L. C. McDermott and E. F. Redish, “Resource Letter: PER-1: Physics Education Research,” Am. J. Phys. **67**(9), 755–767 (1999).

[4] E. E. Prather, T. F. Slater, and E. G. Offerdahl, “Hints of a fundamental misconception in cosmology,” Astr. Ed. Rev. **1**(2), 28–34 (2002).

[5] E. E. Prather, “Students’ Beliefs about the Role of Atoms in Radioactive Decay and Half-life,” J. Geosci Ed. **53**(4), 345–354 (2005).

[6] J. M. Bailey, “Development of a Concept Inventory to Assess Students’ Understanding and Reasoning Difficulties about the Properties and Formation of Stars,” Ph.D. thesis, The University of Arizona, Dept. of Education, 2006.

[7] E. M. Bardar, E. E. Prather, K. Brecher, and T. F. Slater, “Development and Validation of the Light and Spectroscopy Concept Inventory,” Astr. Ed. Rev. **5**(2), 103–113 (2007).

[8] Richard R. Hake, “Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses,” Am. J. Phys. **66**(1), 64–74 (1998).

[9] D. W. Hudgins, E. E. Prather, D. J. Grayson, and D. P. Smits, “Effectiveness of Collaborative Ranking Tasks on Student Understanding of Key Astronomy Concepts,” Astr. Ed. Rev. **5**(1), 1–22 (2006).

[10] E. Prather, T. Slater, J. Adams, J. M. Bailey, L. V. Jones, and J. A. Dostal, “Research on a Lecture-Tutorial Approach to Teaching Introductory Astronomy for Nonscience Majors,” Astr. Ed. Rev. **3**(2), 122–136 (2004).

[11] A. Lightman and P. Sadler, “Teacher Predictions versus Actual Student Gains,” Phys. Teach. **31**(3), 162–167 (1993).

[12] N. Finkelstein, “Researching Reform: Studies of Transforming Undergraduate Physics (& Other Advances in Physics Education Research),” talk presented at the Weekly University of Arizona Physics Colloquium, Tucson, AZ (2006) (private communication).

[13] P. M. Sadler, “The Initial Knowledge State of High School Astronomy Students,” Ed.D. thesis, Harvard University, 1992.

[14] E. F. Reddish, J. M. Saul, and R. N. Steinberg, “Student expectations in introductory physics,” Am. J. Phys. **66**(3), 212–224 (1998).

[15] E. Seymour and N. M. Hewitt, *Talking about Leaving: Why Undergraduates Leave the Sciences*, (Westview, Boulder, CO, 1997).

[16] M. Zeilik and V. J. Morris, “An Examination of Misconceptions in an Astronomy Course for Science, Mathematics, and Engineering Majors,” Astr. Ed. Rev. **2**(1), 101–119 (2003).

[17] M. Zeilik, C. Schau, and N. Mattern, “Conceptual Astronomy. II. Replicating Conceptual Gains, Probing Attitude Changes Across Three Semesters,” Am. J. Phys. **67**(10), 923–927 (1999).

[18] G. Brissenden, T. F. Slater, R. Mathieu, and NISE College Level-One Team, “The Role of Assessment in the Development of the College Introductory Astronomy Course: A “How-To” Guide for Instructors,” Astr. Ed. Rev. **1**(1), 1–24 (2002).

[19] *How People Learn: Brain, Mind, Experience, & School*, edited by J. D. Bransford, A. L. Brown, and R. R. Cocking (National Academies, Washington, DC, 1999).

[20] A. Fraknoi, “Enrollments in Astronomy 101 Courses: An Update,” Astr. Ed. Rev. **1**(1), 121–123 (2002).

[21] F. Lawrenz, D. Huffman, and K. Appeldoorn, “Enhancing the Instructional Environment: Optimal Learning in Introductory Science Classes,” J. Coll. Sci. Teach. **34**(7), 40–44 (2005).

[22] R. S. Lindell, “Enhancing College Students’ Understanding of Lunar Phases,” Ph.D. thesis, Dept. of Physics, The University of Nebraska—Lincoln, 2002.

[23] J. M. Keller, “Development of a Concept Inventory Addressing Students’ Beliefs and Reasoning Difficulties Regarding the Greenhouse Effect,” Ph.D. thesis, Department of Lunar and Planetary Sciences, The University of Arizona, 2006.

[24] E. M. Bardar, E. E. Prather, K. Brecher, and T. F. Slater, “The Need for a Light and Spectroscopy Concept Inventory for Assessing Innovations in Introductory Astronomy Survey Courses,” Astr. Ed. Rev. **4**(2), 20–27 (2006).

[25] T. F. Slater, J. P. Adams, G. Brissenden, and D. Duncan, “What topics are taught in introductory astronomy courses?” Phys. Teach. **39**(1), 52–55 (2001).

[26] M. Zeilik and V. J. Morris-Dueer, “What are essential concepts in ‘Astronomy 101’? A new approach to find consensus from two different samples of instructors,” Astr. Ed. Rev. **3**(2), 61–108 (2005).

[27] L. S. Shulman, “Those Who Understand: Knowledge Growth in Teaching,” Educ. Res. **15**(2), 4–14 (1986).

[28] L. S. Shulman, “Knowledge and Teaching: Foundations of the New Reform,” Harv. Educ. Rev. **57**(1), 1–22 (1987).

[29] *Examining Pedagogical Content Knowledge: The Construct & its Implications for Science Education*, edited by J. Gess-Newsome and N. G. Lederman (Kluwer Academic, Norwell, MA, 1999).

[30] S. Magnusson, J. Krajck, and H. Borko, “Nature, Sources and Development of Pedagogical Content Knowledge for Science Teaching,” in *Examining Pedagogical Content Knowledge: The Construct & its Implications for Science Education*, edited by J. Gess-Newsome and N. G. Lederman (Kluwer Academic, Norwell, MA, 1999).

[31] E. M. Bardar, “First Results from the Light and Spectroscopy Concept Inventory,” Astr. Ed. Rev. **6**(2), 75–84 (2008).

[32] G. Deming and B. Hufnagel, “Who’s taking ASTRO 101?” Phys. Teach. **39**, 368–369 (2001).

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

[34] K. Cummings, J. Marx, R. Thornton, and D. Kuhl, “Evaluating innovation in studio physics,” Am. J. Phys. **67**(7), S38–S44 (1999).

[35] P. A. Kraus, “Promoting Active Learning in Lecture-Based Courses: Demonstrations, Tutorials, and Interactive Tutorial Lectures,” Ph.D. dissertation, Department of Physics and Astronomy, University of Washington, 1997, University Microfilms, UMI No. 9736313.

[36] The Center for Astronomy Education (CAE) academic listserv for astronomy teaching and learning is called Astrolrner@CAE and can be found on the web at http://astronomy101.jpl.nasa.gov/discussion.