

## Comparison of performance on multiple-choice questions and open-ended questions in an introductory astronomy laboratory

Michelle M. Wooten,<sup>1,\*</sup> Adrienne M. Cool,<sup>1</sup> Edward E. Prather,<sup>2</sup> and Kimberly D. Tanner<sup>3</sup>

<sup>1</sup>*Department of Physics and Astronomy, San Francisco State University,  
San Francisco, California 94132, USA*

<sup>2</sup>*Department of Astronomy and Stewart Observatory, University of Arizona, Tucson, Arizona 85721, USA*

<sup>3</sup>*Department of Biology, San Francisco State University, San Francisco, California 94132, USA*

(Received 1 November 2013; published 14 July 2014)

When considering the variety of questions that can be used to measure students' learning, instructors may choose to use multiple-choice questions, which are easier to score than responses to open-ended questions. However, by design, analyses of multiple-choice responses cannot describe all of students' understanding. One method that can be used to learn more about students' learning is the analysis of the open-ended responses students' provide when explaining their multiple-choice response. In this study, we examined the extent to which introductory astronomy students' performance on multiple-choice questions was comparable to their ability to provide evidence when asked to respond to an open-ended question. We quantified students' open-ended responses by developing rubrics that allowed us to score the amount of relevant evidence students' provided. A *minimum rubric score* was determined for each question based on two astronomy educators' perception of the minimum amount of evidence needed to substantiate a scientifically accurate multiple-choice response. The percentage of students meeting both criteria of (1) attaining the minimum rubric score and (2) selecting the correct multiple-choice response was examined at three different phases of instruction: directly before lab instruction, directly after lab instruction, and at the end of the semester. Results suggested that a greater proportion of students were able to choose the correct multiple-choice response than were able to provide responses that attained the minimum rubric score at both the post-lab and post-instruction phases.

DOI: [10.1103/PhysRevSTPER.10.020103](https://doi.org/10.1103/PhysRevSTPER.10.020103)

PACS numbers: 01.40.-d

### I. INTRODUCTION

#### A. Background and motivation

When instructors seek to assess their students' knowledge, a variety of methods may be employed including multiple-choice questions, open-ended questions, projects, written reports, and presentations. Each of these methods has the potential to reveal components of students' abilities and discipline-based knowledge. Multiple-choice questions are commonly used because they allow instructors to quickly quantify varying degrees of knowledge in their classroom and determine what students did or did not gain from instruction. Despite the ease with which multiple-choice questions can be analyzed, the practice of depending on such questions to provide the only evidence regarding students' understanding warrants special attention and consideration for the science education community. Students' abilities to perform well on different types of questions such as those that ask them to support or dispute a

claim using relevant evidence may be missed if multiple-choice questions are used as our sole or primary indicator.

Our study seeks to compare introductory astronomy students' performance on related multiple-choice and open-ended questions. Several studies have shown that students' abilities to perform well on multiple-choice questions have no significant relationship with their ability to perform well on essay tests [1–3]. Furthermore, in the discipline of astronomy education research, two studies revealed that when students were asked to provide explanations of the reasoning behind their multiple-choice responses, the majority of students provided insufficient evidence to match their correct multiple-choice response before and after instructional interventions [4,5]. A similar science education study comparing responses by Lee, Liu, and Linn showed that results from open-ended responses were more useful in determining the range and impact of an instructional intervention [6]. In our study, we used similar methods of comparing multiple-choice and open-ended responses as these studies. We focused on the content area of celestial motions, investigating both similar and different concepts than those previously studied [4]. Unlike any of the other studies mentioned here, we also measured student performance at the end of the semester, in addition to before and after instruction. Comparing students' multiple-choice and open-ended responses at the end of the semester, after

\* [mwootenastronomy@gmail.com](mailto:mwootenastronomy@gmail.com)

an extended period of time had passed since instruction, allowed us to examine whether students retained or changed in their abilities to perform on each type of question.

In this study, we investigated the extent to which students could provide a minimal amount of evidence to warrant their scientifically accurate multiple-choice responses. To do this we quantified students' open-ended responses based on rubrics we developed. A *minimum rubric score* was determined for each question based on two astronomy educators' perception of the minimal evidence needed to substantiate a scientifically accurate multiple-choice response. A comparison of the percentages of students both attaining the minimum rubric score and choosing the correct multiple-choice answer was performed at three different phases of instruction (PRE-LAB, POST-LAB, and POST-COURSE). Analysis of the results provides evidence that introductory astronomy laboratory students' ability to provide a correct multiple-choice response may overestimate their ability to provide minimal evidence for their response.

The assessments and rubrics used for analysis are provided for instructors and researchers interested in either using our rubrics for classroom or research analysis or as an aid to developing their own rubrics for open-ended response analysis.

### B. Participant population

The students participating in this study were mostly nonscience majors at a large urban four-year university attempting to satisfy a general education laboratory requirement. They were enrolled in one of M.W.'s two introductory astronomy laboratory sections during the Fall 2008 semester. A demographic survey was used during the first week of classes to identify the age, ethnicity, and experience of the students participating in the study. The following statistics are the results of the returned demographics survey ( $N = 34$ ), which was optional for students participating in the study: regarding gender, there were 19 males (56%) and 15 females (44%). The majority of the students (82%) were in the age range of 18–22 years. A question about ethnicities revealed a study population of White and Non-Latino (43%), Asian, Pacific Islander, or Filipino (27%), Latino, Chicano, or Mexican American (19%), African American (3%), with the remaining undeclared. Out of the four class standings (freshman, sophomore, junior, senior), most were sophomores (47%) with an almost equal spread across the other class standings. Almost half of the students (44%) were currently enrolled in the lecture counterpart to the course, while one-third (33%) had taken the lecture course the prior year. When asked about students' prior experience in astronomy, a small portion of students reported taking part in prior astronomy related activities (less than 20%).

### C. Course description

The introductory astronomy laboratory that was the focus of this study has been taught to nonscience majors for over two decades. The course description for the laboratory is "Fundamentals of astronomical observation, including optics and spectroscopy. Planetarium exploration of stars, Sun, and Moon. Opportunity for telescopic observation." Students enrolled in the astronomy lab must have taken either the introductory astronomy lecture course offered by the same university or an equivalent at another university or be concurrently enrolled in the lecture course. The laboratory is taught once a week and is two hours and forty-five minutes in duration. Typical enrollment is 30 students. University physics and astronomy graduate students with an interest in astronomy are the usual instructors for the course.

Data presented here were collected in the context of a new astronomy laboratory curriculum based on the 5E model (engage, explore, explain, elaborate, and evaluate) of conceptual change [7,8]. The questions that drive each stage of the 5E model and example activities for one of the labs are presented in Appendix A. To begin designing the curriculum we used the educational theory of backward design—designing the learning objectives and assessments before designing the curriculum [9]. Considering the laboratory course description, two overarching lab objectives were established to address the teaching of celestial motions: Students will be able to (1) find and describe the location of celestial objects using compass direction, altitude, angles between stars, and magnitude and (2) predict the apparent motions of the Sun, Moon, and stars from different positions on Earth. These overarching objectives were broken into smaller learning objectives for five labs. The learning objectives per lab may be found in Appendix B.

## II. ASSESSMENT DESIGN AND ADMINISTRATION

### A. Nature of the assessments

Assessments were chosen or developed with the goal of detecting a deep level of conceptual understanding. Each assessment consisted of a multiple-choice question followed by an open-ended question which asked students to explain in detail the reasoning for their response to the multiple-choice question. Students answered both questions at the same time on each of the assessments. Whenever possible, we used published questions from the Astronomy Diagnostics Test [10], Paul Green's *Peer Instruction for Astronomy* [11], and the Center for Astronomy Education assessment question bank. For those concepts which there were no published assessments, we collaborated to design questions and piloted the questions in the classroom in an effort to establish construct validity. One style of question we developed, not found in literature, made use of a challenge statement, written in everyday language, which conveyed a truism, misconception, or

conceptually rich context. The statement was followed by a Likert scale (strongly disagree, disagree, agree, strongly agree, and don't know) upon which students were asked to rate their level of agreement. This style of question is referred to as a Challenge Statement.

### B. Assessment administration timeline

For each laboratory, approximately three content-oriented assessments were asked to gauge student understanding: a total of 15 content-related assessments were asked over the course of the 5 laboratories. Assessments containing challenge statements were typically asked first so that students would be challenged to reflect on their general knowledge and apply it before being asked about more specific understanding, usually evoked by the assessments containing conceptually rich multiple-choice questions.

Each of the five laboratories used in this study, laboratories 1–5, took approximately two laboratory sessions. The PRE-LAB assessments for Lab 1 were given before Lab 1, and the POST-LAB assessments for Lab 1, which were the same questions as the PRE-LAB assessments, were given directly after Lab 1 instruction. Students were not handed back their responses to the PRE-LAB assessments until after the POST-LAB assessments had taken place. When being graded for credit in the course, PRE-LAB assessments were scored based on the amount of thought provided, not accuracy. POST-LAB assessments were scored on a scale of 0 to 3: A person attaining a 3 provided a scientifically accurate response. Correct responses were never marked on graded work, nor mentioned in class unless students' asked for clarification.

The same assessment strategy was used for labs 2–5. The remainder of the astronomy laboratory instruction was based on curriculum historically used and took approximately four additional weeks. The final exam (POST-COURSE) included all the same assessments that were given as PRE-LAB and POST-LAB assessments from laboratories 1–5 combined.

Table I illustrates the duration of each laboratory module and when the PRE-LAB, POST-LAB, and POST-COURSE assessments were administered.

### C. Method of choosing assessments for analysis

Of the original 15 content-oriented assessments, 7 were chosen for analysis and are presented in the Appendices C—I. The 8 assessments not chosen for analysis were eliminated for one of the two following reasons:

- (1) Despite having five minutes to respond to each assessment, the study population as a whole did not write enough for useful analysis.
- (2) Wording in the question was confusing to a substantial portion of students' such that analysis would not produce meaningful results.

The learning objectives addressed in the 7 assessments are presented here:

- (i) Given the position of Polaris, predicts which stars rise and set. Predicts how stars move with respect to Polaris over the course of 24 hours.
- (ii) Predicts changes in the Sun's maximum altitude over the course of a year
- (iii) Explains how the visible portion of the celestial sphere changes depending on latitude
- (iv) Conceptualizes the role of the Earth's orbit in creating changes in view
- (v) Predicts daily and monthly changes of the Sun's location along the ecliptic using the Zodiacal constellations as a reference
- (vi) Explains why we see the Moon go through phases
- (vii) Given a Moon phase, predicts the phase of the Moon a given number of days later or earlier

Of the 7 assessments used in this research, two are in the form of Challenge Statements and five contain conceptually rich multiple-choice questions, four of which were adapted from published literature. The responses to these assessments comprise the data for the analyses presented in this paper.

## III. ANALYSIS OF STUDENTS' MULTIPLE-CHOICE AND OPEN-ENDED RESPONSES

### A. Analysis of multiple-choice responses

To assess conceptual change in students based on their multiple-choice responses, students were categorized according to whether they choose the correct response or not at each phase of instruction for each assessment. The McNemar Test for Significance of Changes was used to

TABLE I. Assessment time line.

Week	Assessment administration		
1	Introduce course. Administer consent forms and demographics survey.		
2 and 3	PRE-LAB 1 assessments	LAB 1	POST-LAB 1 assessments
4 and 5	PRE-LAB 2 assessments	LAB 2	POST-LAB 2 assessments
6 and 7	PRE-LAB 3 assessments	LAB 3	POST-LAB 3 assessments
7 and 8	PRE-LAB 4 assessments	LAB 4	POST-LAB 4 assessments
9	PRE-LAB 5 assessments	LAB 5	POST-LAB 5 assessments
10–13	Traditional lab instruction		
14	Final Exam: All assessments administered POST-COURSE		

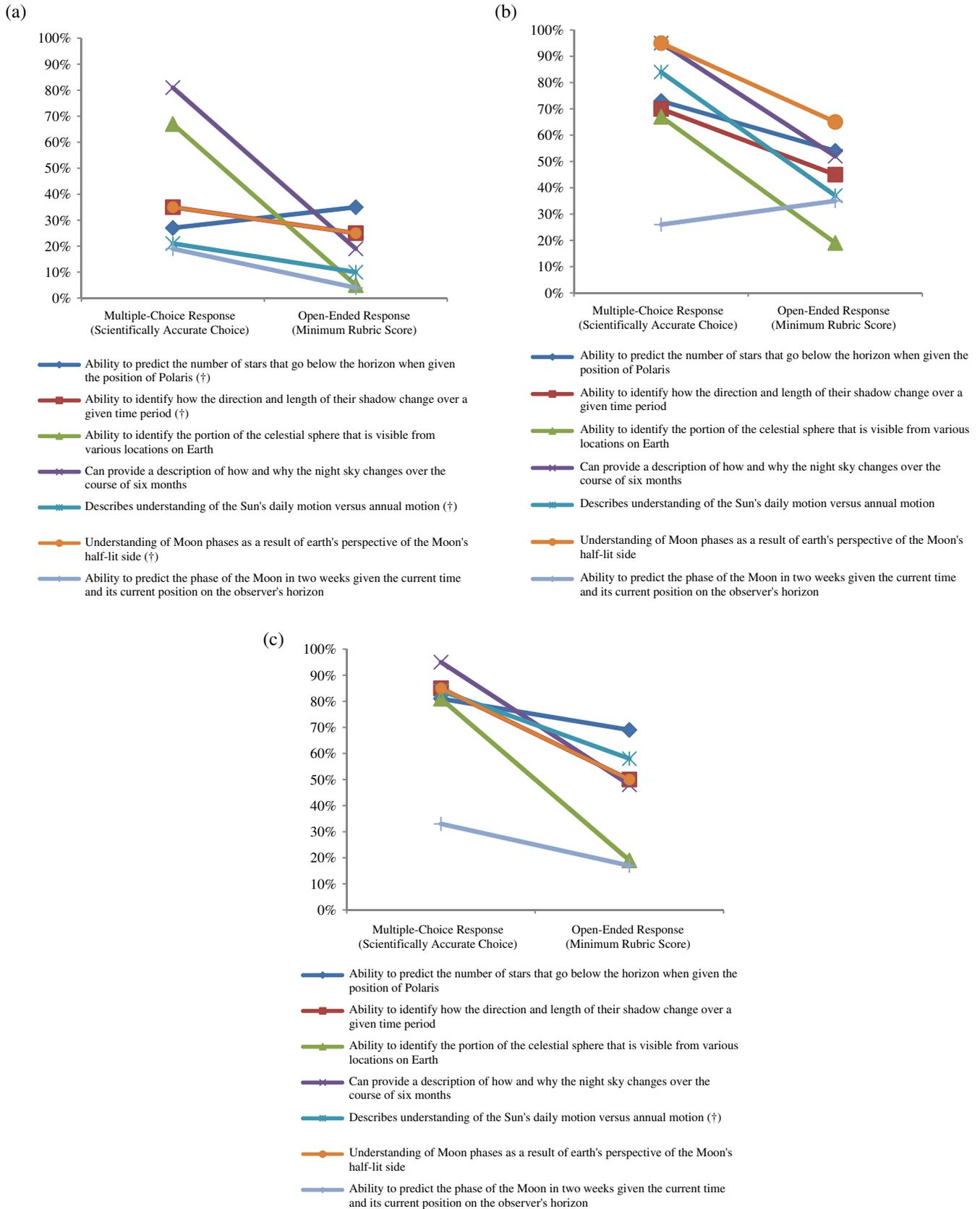


FIG. 1. Percentage of students meeting the minimum criteria (a) PRE-LAB, (b) POST-LAB, and (c) POST-COURSE. We denote when students choosing the correct multiple-choice response are statistically related to the students attaining the minimum rubric score with a dagger.

TABLE II. Percentage of students meeting both the minimum criteria. The asterisks indicate when there was a statistically significant change in the proportion of students meeting the minimum criteria PRE-LAB compared to POST-LAB or POST-INSTRUCTION.

	n = 26			n = 20			n = 21			n = 21			n = 19			n = 20			n = 23		
	PRE-LAB	POST-LAB	POST-COURSE																		
Ability to predict the number of stars that go below the horizon when given the position of Polaris	27%	73%*	81%*	35%	70%	70%	35%	70%	70%	35%	70%	70%	35%	70%	70%	35%	70%	70%	35%	70%	70%
Ability to identify how the direction and length of their shadow change over a given time period				85%*			85%*			85%*			85%*			85%*			85%*		
Ability to identify the portion of the celestial sphere that is visible from various locations on Earth				67%	67%	67%	67%	67%	67%	67%	67%	67%	67%	67%	67%	67%	67%	67%	67%	67%	67%
Can provide a description of how and why the night sky changes over the course of six months				95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%	95%
Describing understanding of the Sun's daily motion versus annual motion				21%	84%*	84%*	21%	84%*	84%*	21%	84%*	84%*	21%	84%*	84%*	21%	84%*	84%*	21%	84%*	84%*
Understanding of Moon phases as a result of Earth's perspective of the Moon's half-lit side				35%	95%*	95%*	35%	95%*	95%*	35%	95%*	95%*	35%	95%*	95%*	35%	95%*	95%*	35%	95%*	95%*
Ability to predict the phase of the Moon in two weeks given the current time and its current position on the observer's horizon				19%	26%	26%	19%	26%	26%	19%	26%	26%	19%	26%	26%	19%	26%	26%	19%	26%	26%
Multiple-Choice Response (Scientifically Accurate Choice)	35%	54%	69%*	50%	45%	45%	50%	45%	45%	50%	45%	45%	50%	45%	45%	50%	45%	45%	50%	45%	45%
Open-Ended Response (Minimum Rubric Score)																					

assess whether there was a statistical difference between students who chose the correct multiple-choice responses PRE-LAB compared to POST-LAB, as well as PRE-LAB compared to POST-COURSE. The null hypothesis of the McNemar Test is that students' ability to choose a correct multiple-choice response will not change as a result of instruction. Therefore, attaining statistical significance ( $p < 0.05$ ) indicates that the observed changes due to instruction differed significantly from those changes expected under the null hypothesis by more than could be attributed to chance.

**B. Analysis of open-ended responses**

Similar to analyses used by Hudgins *et al.* and Wallace, Prather, and Duncan, rubrics were used to analyze students' written responses to the open-ended assessment prompts [4,5]. For each assessment, we developed conceptual categories of ideas that would be expected in a scientifically accurate response to the assessment. If students included ideas from one of these categories in their response, they received a point for that conceptual category in the rubric. Therefore, rubrics allowed us to represent students' responses as a numeric score. In general, rubrics were designed for two purposes: (1) to capture the range of ideas and concepts offered in students' responses, and (2) to estimate students' abilities to offer multiple lines of reasoning that demonstrated understanding of the concept.

To quantify conceptual change based on the open-ended responses using the rubric scores, a *minimum rubric score* was determined for each assessment. The *minimum rubric score* represented the number of rubric categories a student would need to include in their response for it to be considered representative of a minimal understanding of the concept. The value of the minimum rubric score was discussed and agreed upon for each rubric by two researchers with astronomy backgrounds. We note that a student could therefore represent correct ideas in their response and still not attain the minimum rubric score.

Rubrics and the minimum rubric score for each assessment are provided in the Appendices beneath their respective assessment.

To attain interrater reliability, 10% of students' responses were randomly selected from the pool of PRE-LAB, POST-LAB, and POST-COURSE responses for each of the assessments. Two of us then used the rubrics to independently score each student response. An overall Cohen's kappa coefficient of 0.75 or higher was obtained for 15 out of the 25 possible rubric categories. For the 10 categories that attained a lower Cohen's Kappa coefficient than 0.75, we discussed discrepancies in their ratings and refined the rubrics accordingly. M. W. then proceeded to score the entirety of the student responses based on the refined rubrics.

As with the analysis of students' multiple-choice responses, students were categorized as to whether or

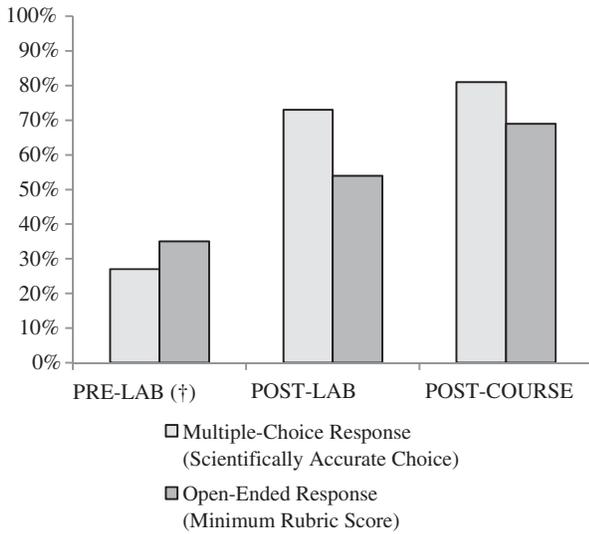


FIG. 2. Percentage of students meeting minimum criteria when asked about the apparent motion of stars near Polaris with respect to the horizon ( $n = 26$ ).

not they attained the minimum rubric at the three phases of instruction. The McNemar test for significance of changes was used to assess whether there was a statistical difference ( $p < 0.05$ ) in the proportion of students who attained a minimum rubric score PRE-LAB compared to POST-LAB, as well as PRE-LAB compared to POST-COURSE.

**C. Statistical analyses performed to compare multiple-choice and open-ended responses**

The percentage of students correctly responding to the multiple-choice question on an assessment was plotted and compared to the percentage of students attaining the minimum rubric score from their open-ended response on the same assessment. A chi-square test for independence was used to determine whether an individual student’s

ability to correctly respond to a multiple-choice question was related to their ability to attain a minimum rubric score for each assessment at each phase of assessment. The null hypothesis of the chi-square test for independence is that a student’s ability to correctly respond to a multiple-choice question is independent of their ability to attain a minimum rubric score on the same assessment. Therefore, if a significance value of  $p < 0.05$  was attained, the null hypothesis was rejected: a student’s ability to correctly respond to the multiple-choice question was statistically related to their ability to attain the minimum rubric score. This test was performed for each of the seven assessments at each phase of instruction (PRE-LAB, POST-LAB, and POST-COURSE). A *post hoc* power analysis was conducted with the program G\* POWER 3 to determine the power associated with each chi-square test for independence performed at each phase of every assessment [12]. Power indicates the probability that the null hypothesis is correctly rejected in order to avoid false positive results. Power analysis for the chi-square test for independence is dependent on the sample size, the significance level ( $\alpha = 0.05$ ), and the effect size. The effect size was calculated using the following:

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

Students who both chose the correct multiple-choice response and attained a minimum rubric score for their response to each assessment are referred to in the analysis as having “met both minimum criteria.”

**IV. RESULTS OF ASSESSMENT ANALYSIS**

**A. Combined assessments results**

We begin our presentation of results by examining any overarching trends in students’ ability to choose correct

TABLE III. Example open-ended responses given for the assessment regarding the apparent motion of stars near Polaris with respect to the horizon.

Minimum criteria met	Example open-ended responses
Correct multiple-choice <i>and</i> minimum rubric score	Stars that are at a greater distance from Polaris than Polaris is from the horizon will dip below the horizon. The two stars closer to Polaris will be circumpolar stars. The Earth’s rotation on its axis creates the illusion that celestial bodies are making a circle around us, while Polaris barely moves because it is very near the celestial pole. As the 3 more distant stars circle Polaris, their paths will carry them below the horizon. <i>Student J, POST-COURSE</i>
Correct multiple-choice response only	The distance between Polaris and the outermost three stars exceeds the distance between Polaris and the horizon, so they will rotate to a point below the horizon, for each of the three. <i>Student N, POST-LAB</i> The three stars that are the farthest away from Polaris will at some time over the next 24 hours will dip below the Horizon. For them to make a complete rotation around Polaris they have to go below the horizon. <i>Student HH, POST-LAB</i>

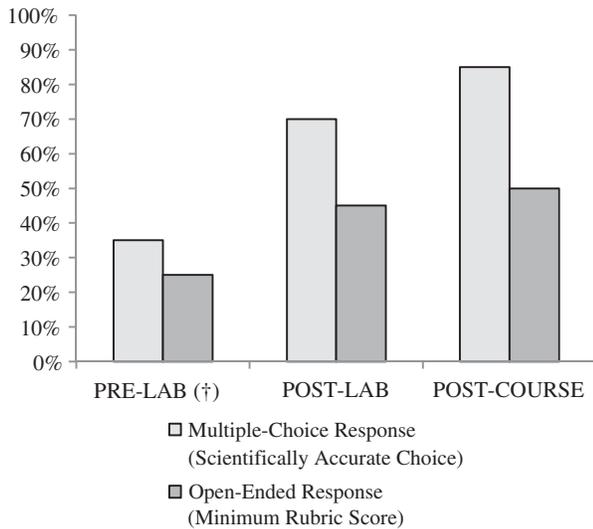


FIG. 3. Percentage of students meeting minimum criteria when asked about the length and direction of their shadow over a two week period ( $n = 20$ ).

multiple-choice responses and attain a minimum rubric score. Figure 1 is a graphical representation of the percentages of students who have met either of the minimum criteria at each phase of assessment. We have designated when a statistical relationship exists between a student’s ability to meet both of the minimum criteria with a dagger. The trend that students were more proficient in choosing correct multiple-choice responses than providing minimum evidence to support their response at all phases of instruction is apparent in this figure.

The percentages of students meeting each of the minimum criteria for each assessment are provided in numerical form in Table II. The statistical significance of the changes

in the proportion of students able to meet either of the minimum criteria from PRE-LAB to POST-LAB and PRE-LAB to POST-COURSE are labeled in the respective POST phase percentage with an asterisk. For instance, for the assessment gauging students’ “Ability to predict the number of stars that go below the horizon when given the position of Polaris,” the change in students’ ability to choose a correct multiple-choice response from PRE-LAB to POST-LAB and PRE-LAB to POST-COURSE are both statistically significant. Upon examining this Table, there is a trend that suggests that students improve in their performance on both minimum criteria from before to after instruction. However, when comparing the results from one assessment to another, there is no apparent trend in the extent to which students improved on their multiple-choice response performance compared to their open-ended response performance. Therefore, we find it constructive to examine students’ responses to each question, individually.

**B. Individual assessment results**

To compare students’ multiple-choice responses with their ability to provide minimal reasoning on each assessment, we plotted the percentage of students attaining each of the minimum criteria PRE-LAB, POST-LAB, and POST-COURSE in two forms side by side: (1) the percentage of those who chose the correct multiple-choice response and (2) the percentage of those who obtained a minimum rubric score. When there is a relationship between a student’s ability to meet both the minimum criteria, we have denoted a dagger by that phase of assessment.

We present the results of the analysis of each of the seven assessments in the same order that they are presented in Fig. 1 and in Table II. During the presentation of the

TABLE IV. Example open-ended responses given for the assessment about the changing length and direction of a shadow over a two week period.

Minimum criteria met	Example open-ended responses
Correct multiple-choice <i>and</i> minimum rubric score	The Sun will be lower in the south giving your shadow the appearance of being longer. I don’t exactly know how to explain the length of the shadow in words and my diagram will hopefully help [diagram shows sun in south and person’s shadow pointing north]. The sun is lower and hits more of your body giving the shadow more length. <i>Student I, PRE-LAB</i>
Correct multiple-choice response only	My shadow will be longer and pointing toward the north because at this time of year, our days get shorter and the sun begins to rise slightly lower. This causes us to receive sunlight at an angle, which upon contact creates a longer shadow. The shadow will be pointing north because the sun, from our location, will be at its highest point at noon in the southern sky. This is also after the summer solstice. <i>Student LL, POST-LAB</i> Because as the Sun is going south toward the winter solstice. And from this duration the Sun’s light is deepening. So our shadows will be longer coming from the direction south pointing toward the north. <i>Student B, POST-COURSE</i>

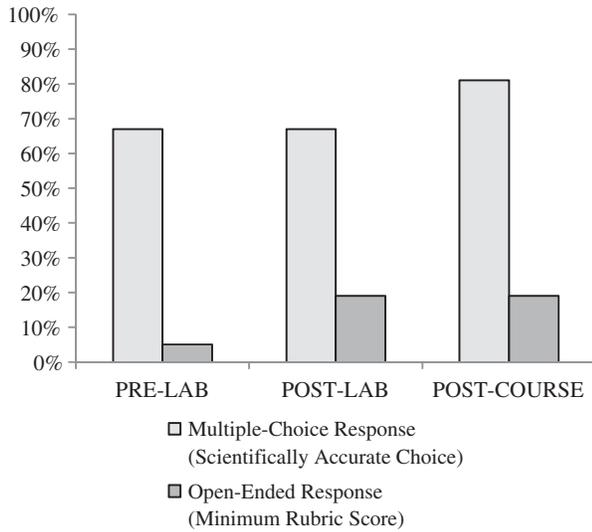


FIG. 4. Percentage of students meeting minimum criteria when asked about the portion of the celestial sphere visible from different latitudes ( $n = 21$ ).

analysis, we provide example responses given by students to familiarize the reader with the type of evidence that students needed to provide to attain the minimum rubric score.

**1. Students’ ability to predict the number of stars that go below the horizon when given the position of Polaris**

Figure 2 shows the distributions of the percentage of students meeting each of the minimum criteria when asked to predict the number of stars that go below the horizon when given the position of Polaris (Fig. 2; see Appendix C). These distributions reveal a pattern one might expect to see when students learn as a result of instruction: the growth in the number of students able to choose the correct multiple-choice response and correctly reason about their response increases after instruction and then again by the end of the semester. Statistical analysis revealed a relationship between a student’s ability to meet both the minimum criteria PRE-LAB ( $\chi^2 = 18.1$ ) at a power of 0.98. However, there was not enough evidence to support a relationship between a student’s ability to meet both the minimum criteria at the POST-LAB and POST-COURSE phases of instruction. We discuss this phenomenon further in the Study Limitations portion of the Discussion section.

Table III conveys the example open-ended responses that students gave when asked to explain their reasoning regarding their multiple-choice response in this assessment (see Appendix C). The responses are divided based on

TABLE V. Example open-ended responses given for the assessment about the portion of the celestial sphere visible from different latitudes.

Minimum criteria met	Example open-ended responses
Correct multiple-choice <i>and</i> minimum rubric score	I say this because at the north pole, all the stars are circumpolar, which would mean that we’d never see a different pattern of stars. At a latitude as low as ours, we are able to see quite an array of stars, stars that come from 52 degrees south even. We do see some circumpolar stars, but many are rise and set stars. Now, at the equator, ALL of the stars are rise and set stars. There are no circumpolar stars, so we never see the same stars year round. We can see stars that come from the south at certain times of the year, as well as stars that come from the north. We would be able to see a great variety of stars, never having to see the same pattern over and over again through the year. <i>Student E, POST-COURSE</i>
Correct multiple-choice response only	To see greatest number of stars throughout the period of one year would be the Equator, because of the tilt the Equator is more capable of seeing a variety of stars. The North Pole is restricted to seeing the same stars that are mostly circumpolar, so that can’t be the answer. SF is located more North and is still limited to a certain amount of stars so that can’t be the answer. And D is definitely not the answer because all locations on Earth do NOT see the same number of stars like I explained about the North Pole and its circumpolar stars. So the Equator is the best answer. The equator is at the center. <i>Student M, POST-COURSE</i>  I have a hard time justifying an answer because every location will have a similar amount of stars in the 1/2 of the celestial sphere they can view. Only some locations have more variety due to the Earth’s tilt. A location near the equator will be able to see stars from both the northern and southern hemisphere, but a location like that with the latitude of San Francisco will always have 1/2 the celestial sphere that includes a north star. Its hard to say if one place has more stars or not, but I guess going with variety which makes sense as having seen more stars would have to be near the equator. <i>Student I, POST-COURSE</i>

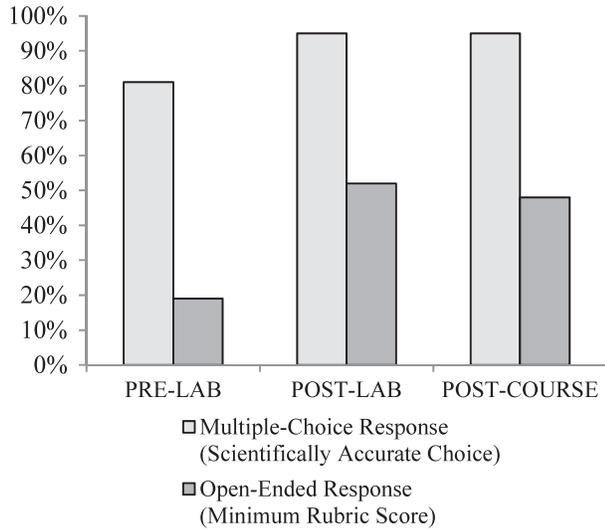


FIG. 5. Percentage of students meeting minimum criteria when asked whether or not the same stars are visible in an observer’s night sky six months apart ( $n = 21$ ).

whether or not the student chose the correct multiple-choice response for the same assessment.

**2. Students’ ability to identify how the direction and length of their shadow change over a given time period**

When asked to identify the length and direction of their shadow over a given time period, a similar trend is seen in

the distribution of responses of students meeting the minimum criteria that were discussed in the previous assessment: there is an upward trend from PRE-LAB to POST-LAB and then POST-LAB to POST-COURSE (Fig. 3). Also, PRE-LAB, there was a relationship between a student’s ability to meet both the minimum criteria ( $\chi^2 = 12.4$ ) at a statistical power of 0.94.

Table IV conveys the open-ended responses that students gave to this assessment (see Appendix D). The responses are divided based on whether or not the student chose the correct multiple-choice response for the same assessment.

**3. Students’ ability to identify the portion of the celestial sphere that is visible from various locations on Earth**

For the next set of distributions for the assessment regarding the portion of the celestial sphere that is visible from various locations on Earth, at every phase of assessment, less than 20% of the class attained a minimum rubric score while the majority of the class chose the correct multiple-choice response (Fig. 4). Also, while there is an increase in the number of students achieving a minimum rubric score from PRE-LAB to POST-LAB, there is no such increase seen for the number of students choosing the correct multiple-choice from PRE-LAB to POST-LAB.

Table V conveys example open-ended responses that students gave to this assessment (see Appendix E). The responses are divided based on whether or not the student chose the correct multiple-choice response for the same assessment.

TABLE VI. Example open-ended responses given for the assessment about the changing appearance of the night sky over the course six months.

Minimum criteria met	Example open-ended responses
Correct multiple-choice <i>and</i> minimum rubric score	As the Earth rotates the Sun we see different parts of the sky. In 6 months we will be on the opposite side of the sun then we are now so we will see a completely different night sky. [has drawn two earths on opposite sides of an orbit around the Sun with the night sides of earth shaded and words on either side saying “We can see all this sky at night” and “we will be able to see this sky in six months” on the part of the sky that the night sides of the earth are facing] <i>Student HH, POST-LAB</i>
Correct multiple-choice response only	I believe it is impossible for us to see the same pattern of stars. Because the stars aren’t always the same the earth is continually rotating and different stars or constellations appear as time goes by. In my illustration I am trying to show that each as the world rotates the world see something slightly different. <i>Student EE, POST-COURSE</i> I disagree because as Earth rotates around the Sun, the Sun would be blocking certain stars that we cannot see in given months. The pattern of stars six months from now would be different due to the fact that it would be summer. <i>Student CC, POST-COURSE</i> As the Earth rotates 1 degree per day over a 365 day period, the pattern of stars visible in August, for example, will change nearly every month and thus, in 6 months, a different pattern will be visible—Hence 12 constellations being visible through the year, approximately 1 per month. <i>Student MM, POST-LAB</i>

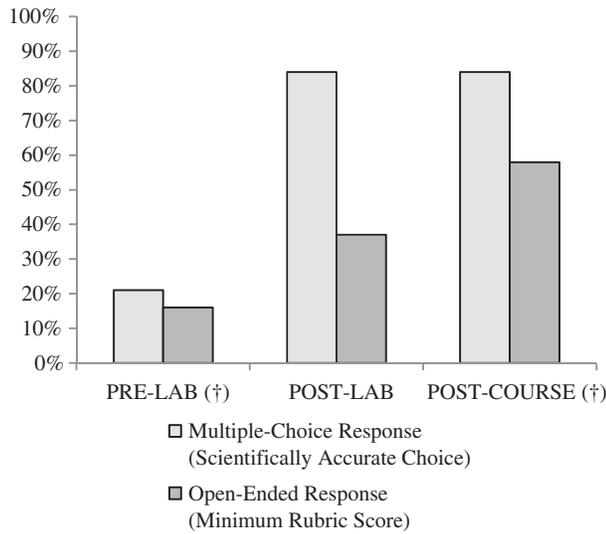


FIG. 6. Percentage of students meeting minimum criteria when asked about the Sun’s apparent motion with respect to the stars over a one day period ( $n = 19$ ).

**4. Students’ ability to provide a description of how and why the night sky changes over the course of sixth months**

In the next set of distributions for the assessment regarding how the night sky changes over the course of six months, the percentage of students choosing correct multiple-choice responses are almost twice as high, or more than the number of students receiving a minimum rubric

score at each phase (Fig. 5). Especially noteworthy is the discrepancy between the percentage of students choosing the correct multiple-choice response PRE-LAB compared to the percentage of students who can provide minimal lines of reasoning to their response PRE-LAB.

Table VI conveys example open-ended responses that students gave to this assessment (see Appendix F). The responses are divided based on whether or not the student chose the correct multiple-choice response for the same assessment.

**5. Students’ ability to describe understanding of the Sun’s daily motion versus annual motion**

For the assessment about the Sun’s apparent motion with respect to the stars over the course of a day, the pattern seen in the percentage of students attaining either of the minimum criteria is representative of the same pattern witnessed in the first two assessments: a general increase from PRE-LAB to POST-LAB and again from POST-LAB to POST-COURSE (Fig. 6). However, one might note that there is a large discrepancy between the percentage of students choosing a multiple-choice response and attaining a minimum rubric score POST-LAB. Unique to all other patterns of student responses in this study, there was a relationship between a student’s ability to meet both the minimum criteria at both the PRE-LAB ( $\chi^2 = 13.4$ ) and POST-COURSE ( $\chi^2 = 4.9$ ) phases of instruction, at a statistical power of 0.95 and 0.6, respectively. This relationship indicates that students have the potential to improve in their ability to provide minimal lines of evidence regarding

TABLE VII. Example open-ended responses given for the assessment regarding the apparent motion of the Sun with respect to the stars over the course of a day.

Minimum criteria met	Example open-ended responses
Correct multiple-choice <i>and</i> minimum rubric score	The Sun would be in front of the same constellation. The daytime doesn’t make that much of a difference. The sun stays in front of the specific constellation for up to more than a month. The stars rise and set, as the Sun does. Only our orbiting the Sun makes a difference on which stars we can see. In order to look at different stars we have to wait for at least a couple of days. Both Sun and stars appear to rise in the east & set in the west. <i>Student D, POST-LAB</i>
Correct multiple-choice response only	The sun would still be in front of Gemini because the zodiac constellation and sun follow the same path called the ecliptic. The sun right now is at noon so in 6 hours the sun is going to be at the horizon. The constellation would be moving as well. Gemini, 6 hours from 12 noon would also be at the horizon making Leo the highest in the sky at 6pm. This is because of the ecliptic plane. <i>Student CC, POST-COURSE</i> When the sun sets the sun will still be in front of the constellation of Gemini because the sun spends approximately one month in each constellation. It would appear as if the sun would set in the west, in front of Pisces, but that is not the case. Therefore, the Sun will still be in front of Gemini when it sets. Though the Sun is highest in the sky at noon, the Sun will still set in front of Gemini. <i>Student H, POST-COURSE</i> The Sun will only shift about 1 degree in relation to the stars behind it. It’s the earth moving around the sun that will change the sun’s positioning in our sky. So it will be in front of Gemini. <i>Student G, POST-LAB</i>

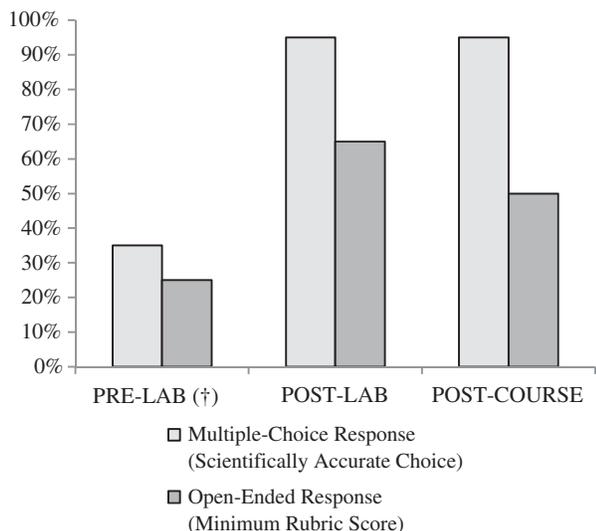


FIG. 7. Percentage of students meeting the minimum criteria when asked about the cause of the Moon’s phases ( $n = 23$ ).

their multiple-choice response considerably by the end of the semester.

Table VII conveys example open-ended responses that students gave to this assessment (see Appendix G). The responses are divided based on whether or not the student chose the correct multiple-choice response for the same assessment.

**6. Students’ understanding of Moon phases as a result of Earth’s perspective of the Moon’s half-lit side**

Figure 7 depicts the distributions of students meeting either minimum criteria when asked about the cause of Moon phases. The ceiling effect represented in the

multiple-choice distribution at the POST-LAB and POST-COURSE distribution is not seen in the minimum rubric score distribution, even though there is a large gain from PRE-LAB to POST-LAB in those meeting either criteria. Also noteworthy is that fewer students achieve a minimum rubric score POST-COURSE than do POST-LAB, perhaps indicating that students do not retain learned evidence to support their multiple-choice response. This is the fourth assessment we have discussed where there is a relationship between a student’s ability to meet both the minimum criteria PRE-LAB ( $\chi^2 = 5.9$ ). The *post hoc* power analysis revealed a power of 0.68 for this phase.

Table VIII conveys example open-ended responses that students gave to this assessment (see Appendix H). The responses are divided based on whether or not the student chose the correct multiple-choice response for the same assessment.

**7. Students’ ability to predict the phase of the Moon in two weeks given the current time and its current position on the observer’s horizon**

The following set of distributions illustrating the percentage of students meeting either of the minimum criteria when asked to predict a future Moon phase from a given time and phase, very few students attained either minimum criterion at any phase (Fig. 8). POST-LAB more students receive a minimum rubric score than do the number of students choosing the correct multiple-choice response. This happens in only one other instance (Fig. 2, PRE-LAB).

Table IX conveys example open-ended responses that students gave to this assessment (see Appendix I). The responses are divided based on whether or not the student chose the correct multiple-choice response for the same assessment.

TABLE VIII. Example open-ended responses given for the assessment regarding the cause of the Moon’s phases.

Minimum criteria met	Example open-ended responses
Correct multiple-choice <i>and</i> minimum rubric score	The reason Earth sees different phases of the moon is not because it blocks the sunlight itself, but because the moon always has 1/2 of its surface lit up, and according to Earth’s position, we only see a section of this light ...the moon appear to go through phases. Moon always 1/2 light up. Only thing that changes is the Earth’s perspective. <i>Student C, POST-LAB</i>
Correct multiple-choice response only	The reason the moon has phases is because it is far away and above the earth enough that the earth does not block it but the sun light is actually hitting the moon. So when we see parts of the moon it’s what the sun is illuminating to allow us to see the moon. <i>Student B, POST-COURSE</i> The moon has its phases because it is always half lit. The sun’s light only hits half of the moon. The phases are seen from the Earth’s perspective. It takes a month to complete a whole phase of new and back to the new phase. The moon has its phases from different locations on its path. As in the drawing you have the Moon’s path, with it being half lit at all times. <i>Student M, POST-COURSE</i>

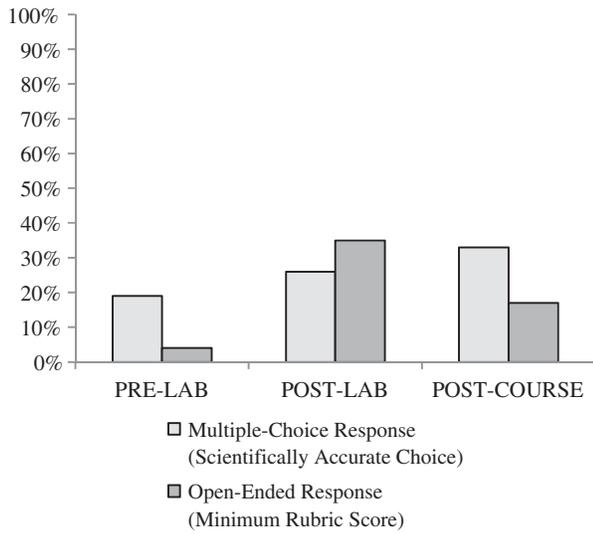


FIG. 8. Percentage of students meeting minimum criteria when asked to predict the phase of the Moon in two weeks given the current time and phase ( $n = 23$ ).

### V. DISCUSSION

This study investigated whether or not introductory astronomy students could provide a sufficient amount of evidence in writing to support their correct multiple-choice responses on assessments that asked them to explain their reasoning for their multiple-choice response. We developed a minimum rubric score, unique to each assessment, to allow us to compare students’ ability to provide minimal lines of reasoning to their ability to provide a correct multiple-choice response. This study is unique in that it examined students’ abilities to meet both minimum criteria (choose the scientifically accurate multiple-choice response and attain the minimum rubric score) in conceptual areas of

celestial motions not yet explored in the literature. Additionally, we compared students’ abilities to meet either minimum criterion at three different phases of lab instruction: before instruction, directly after instruction, and at the end of the semester. This allowed us to see whether or not students’ abilities were changed or retained several weeks after instruction ended. This study is pertinent to astronomy educators and science educators as results from multiple-choice questions are published far more than results from open-ended questions in introductory science courses. However, our findings extend prior research, as we will further detail in this section, that students’ open-ended responses that serve as explanations for their multiple-choice responses are underdeveloped.

To begin our analysis, we will attempt to identify what we would have concluded about student learning had we only studied their responses to multiple-choice questions. Next we take this same approach, instead with respect to the analysis of students’ open-ended responses alone. We then describe what we learn when we consider the results from both students’ multiple-choice response and open-ended responses.

#### A. Conclusions based on the analysis of multiple-choice responses

When considering the results of students’ multiple-choice responses alone, we might infer that students had a desired degree of understanding of the majority of the learning objectives by the time they finished the course. This is evidenced by the observation that at least 80% of students chose the correct multiple-choice response on 6 out of 7 assessments. Additionally, not only did the percentage of students correctly responding to each assessment improve from before instruction to the end of the semester, but for 4 out of 7 assessments, there was at least a

TABLE IX. Example open-ended responses given for the assessment asking students to predict the phase of the Moon two weeks from the current time and Moon phase.

Minimum criteria met	Example open-ended responses
Correct multiple-choice <i>and</i> minimum rubric score	The lunar phase in two weeks would be 180 degrees around its revolution. The initial position would be there because in roughly 3 hours the sun will be visible. The lunar cycle is one month (one full revolution) and two weeks would be on the opposite side of the Earth from its initial position. It would be a waxing gibbous because it is the phase after the new moon. [student’s drawing correctly shows moon at ‘today’s’ position with respect to the earth and Sun and the waxing gibbous position of the moon] <i>Student P, POST-COURSE</i>
Correct multiple-choice response only	In two weeks the moon will be a waning gibbous. Because two weeks ago it was a waxing gibbous. So two weeks from then the moon will be a waning gibbous where more than half the moon illuminated. <i>Student H, POST-LAB</i> The Moon will be a waxing gibbous in two weeks because it will be half way around its circle around earth, put it at waxing gibbous. <i>Student II, POST-COURSE</i>

50% gain in the percentage of students correctly responding to the questions by the end of the semester.

Another conclusion that may be drawn is that when instruction provided a significant increase in knowledge, this knowledge was retained. For example, for all three assessments where statistically significant changes occurred in the percentage of students offering correct responses before instruction to after instruction (PRE-LAB to POST-LAB), the change in the percentage of students choosing correct multiple-choice responses is also significant from before instruction to the end of the semester (PRE-LAB to POST-COURSE). Thus, it appears that students retained learned information when it was first unknown.

A final noteworthy trend in students' performance on multiple-choice questions is that for 6 out of 7 assessments, it appears that students either retained or deepened their knowledge from directly after instruction to the end of the semester (POST-LAB to POST-COURSE). While it is possible that this trend could relate to students responses being returned directly after the POST-LAB phase, correct responses were never marked on the graded work, nor mentioned in class unless students' asked for clarification.

Therefore, from students' multiple-choice responses alone, one might conclude that a majority of students positively increased in their understanding of the learning objectives and retained this understanding until the end of the semester.

### **B. Conclusions based on the analysis of open-ended responses**

In this section, we discuss conclusions that could be made from the percentages of students' providing minimal evidence in their explanations of their multiple-choice response. In general, it would appear that there were only modest improvements in students' ability to provide a minimum amount of evidence in their explanations. For example, there is only one assessment for which the gain in the percentage of students providing minimum evidence was greater than 30%. Additionally, out of all of the assessments, the highest percentage of students providing explanations that met the minimum rubric score was 69%, and this occurred only at a POST-COURSE phase of instruction.

From students' open-ended responses, we might also conclude that it is difficult to teach students in such a way that they gain or retain their ability to articulate key lines of evidence to their responses. In particular, there were three assessments where the increase in the proportion of students attaining the minimum rubric score from before instruction to after instruction (PRE-LAB to POST-LAB) were statistically significant. On only one of these is the increase from before instruction to the end of the course (PRE-LAB to POST-COURSE) significant as well. Further, students' performance on the open-ended responses revealed no apparent trends from post instruction

to the end of the course (POST-LAB to POST-COURSE): On 3 out of 7 assessments, the percentage of students attaining the minimum rubric score increased, on 3 out of 7 the percentage decreased, and for 1 out of 7, the percentage stayed the same. Therefore, it is not clear whether students retained or deepened their understanding of concepts by the end of the semester.

Thus, from students' open-ended responses alone, we might conclude that in general approximately one-quarter of students had a positive increase in their understanding of the learning objectives by the end of the lab, but that the learning was not demonstrably retained at the end of the semester.

### **C. Conclusions based on both multiple-choice and minimum rubric score analysis**

To examine what we might conclude if we were using both students' multiple-choice responses and their open-ended responses as indicators of the depth of their knowledge, we return to Fig. 1. Figure 1 illustrates that more students were able to choose a scientifically accurate multiple-choice response than were able to provide minimum evidence to support their response in 90% (19/21) of the phases assessed. In addition, the results depicted in Table II suggest that statistically significant changes in the proportion of students' multiple-choice responses from PRE-LAB to either the POST-LAB or POST-COURSE phases were not always mirrored by significant changes seen in their open-ended response counterparts.

We considered the possibility that the lowering of the minimum rubric score per assessment might have produced similar results in the open-ended responses to those seen in the multiple-choice responses. For example, if the minimum rubric score was decreased to one for any assessment, more students would attain the minimum rubric score and the percentage of students meeting each of the minimum criteria may appear more equally matched. However, decreasing the minimum rubric score would mean that students would not have to provide one or more strands of evidence that two astronomy educators agreed would be the absolute smallest amount of evidence needed to support a correct multiple-choice response.

### **D. Comparison to previous studies**

Our study's results both support and extend findings by other studies that used different instructional inventions and assessments to measure and compare introductory astronomy students' conceptual knowledge on multiple-choice and open-ended responses.

One such study examined students' pre and post conceptions of cosmology concepts when a collection of lecture tutorials on cosmology that the researchers had developed were used as an instructional intervention [5]. Part of the analyses of the effectiveness of the tutorials compared students' open-ended responses defending their

choice of a graph for the evidence of dark matter. C. W., E. P., and D. D. found that though a striking number of students were able to choose correct closed-ended responses after instruction, students did not provide in their explanations all the reasoning that would warrant a complete explanation. Additionally, it was discovered that with an increase in the complexity of the reasoning required to score highly on the open-ended questions, there was a decrease in the gains for open-ended performance. The researchers questioned whether students' minimal support to their open-ended responses was due to the fact that the students did not receive extra points toward their grade for providing robust responses. Our study, while assessing student understanding of different introductory astronomy concepts than the Wallace, Prather, and Duncan study, provides evidence that even when students receive grade points for providing complete responses, their ability to provide minimal evidence remains underdeveloped [5].

Another related study took place in an introductory astronomy lecture course, where the effectiveness of the instructional intervention called the ranking task was assessed to evaluate the effectiveness on fostering conceptual change [4]. Ranking tasks are tasks where students must order a series of responses in a conceptually rich problem correctly, and provide narrations explaining their ordering. Conceptual change over eight introductory astronomy concepts by the ranking tasks were studied, three of which were relevant to this study: phases of the Moon, motion of the sky, and seasons. The study showed a significant increase in the percent of students answering correctly to conceptually challenging multiple-choice questions after completing the ranking tasks. A portion of students' narrations were analyzed and it was found that while ranking tasks increased the number of students including drawings in their responses as an aid to their explanation, many of these students were not able to provide depth in their explanations that would be characteristic of complete understanding of the concepts. While the Hudgins *et al.* study revealed this misalignment between students' performance on multiple-choice and open-ended responses, the present study provided more depth to this analysis by expanding the number of topics on celestial motions examined, using different types of assessments, and comparing responses at the end of the semester in addition to directly after instruction [4].

### E. Study limitations

Because of the small number of participants in this study, for the majority of the statistical analyses performed to assess whether there was a relationship between a student's ability to meet both the minimum criteria discussed in the Results, "II. Individual assessment results," there was not enough evidence to support the rejection of the null hypothesis. Considering the one-to-one relationship between  $p$  values and power—that a nonsignificant result

is coincident with a low  $p$  value—it is not surprising that the power of these nonsignificant results were all less than 0.5 [13]. Therefore, if more students had been assessed, perhaps we would have found a relationship between a student's ability to meet both the minimum criteria in more of the comparisons than is currently described in this study. An *a priori* power analysis revealed that in order to detect a medium effect size of 0.3 at a power of 0.8 for any given phase of assessment, 88 matching pairs of students' multiple-choice and open-ended responses would be required. For the phases presented in Sec. II that were statistically significant, larger effect sizes (0.6 or greater) were detected, not requiring a greater sample size to determine statistical significance.

### F. Conclusions

Multiple-choice questions are often used to gauge student learning in order to reduce the time that it would take to evaluate learning using other methods [1,14,15]. Such reasons are understandable, especially considering many introductory science courses serve as general education courses and yield high enrollment.

Our comparisons of students' performance on multiple-choice questions to their open-ended responses indicate that interpretations based on students' ability to choose correct multiple-choice responses may serve as an overestimation of their understanding and learning. Indeed, by design, students' responses to multiple-choice questions could not fully describe to us all of their knowledge. When asked to reason about their multiple-choice response, students do not generally provide minimal lines of evidence to support their correct response, perhaps because such open-ended questions rely on students' ability to transfer their thoughts into writing. This may indicate that students do not have a good understanding of the content or their own metacognition regarding the concepts, including why certain lines of evidence are important in supporting their response.

We also suggest that since there was a statistical relationship between a student's ability to both choose the correct multiple-choice responses and provide minimal lines of evidence to support their response before instruction, in general, open-ended responses are a good approximation for the amount of knowledge students hold at any phase of instruction. Based on the evidence presented, we suggest that students' open-ended responses serve as a better gauge of their learning than their multiple-choice responses, as open-ended responses do not appear to overestimate students' understanding in the same way that their multiple-choice responses may. However, because of the different insights multiple-choice and open-ended questions offer, we also suggest that neither measure alone is satisfactory for the evaluation of student learning. Furthermore, we suggest that if we care about our students' ability to explain their reasoning, it seems imperative that we carefully measure their ability to do so.

## ACKNOWLEDGMENTS

We thank the department chairs of the Department of Physics & Astronomy of San Francisco State University for supporting the education-oriented Master's thesis research described in this paper. We thank T. Slater and S. Slater for providing guidance in the development of the learning objectives of the curriculum. We also thank the anonymous referees of this paper for their careful review and suggestions. This work was funded in part by a SFSU Center for Science and Mathematics Education grant.

## APPENDIX A

In Appendix A, a table is presented demonstrating the 5E model phases with examples of questions guiding each phase. Example activities used for each phase during Lab 2 of the introductory astronomy laboratory curriculum described in this study are also presented for clarification.

TABLE X. The 5E model phases with examples of questions guiding each phase, and example activities used for each phase during Lab 2 of the introductory astronomy laboratory curriculum described in this study.

5E Phase	Questions driving phase	Description of instruction used in Lab 2
Engage	What about this topic relates to my students lives? What demonstration can I give my students that match their preconception knowledge level and would stump them?	The instructor revealed pictures of star trails at different locations on Earth and asked students to explain why they think the trails look differently.
Explore	What activity can I give my students that will allow them to explore their thoughts and put together puzzle pieces that lead to a better picture of this concept?	The instructor asked students to explain differences in view while rotating in a chair, looking out versus looking straight above. Afterward, she asked students to compare these views to planetarium views of the stars apparent motion above the horizon from the North Pole and the Equator. Students were asked to describe characteristics of each and draw on plastic domes what they saw.
Explain	What do my students need to hear that will help them put together the pieces explored in the Engage and Explore portions of the lab? What language do they need to gain so that they can talk about this topic using the language of science?	The instructor completed this phase by making and describing a 2D representation of a celestial sphere on the board. She drew the horizon line of a person at each of the latitudes explored and labeled key components of the celestial sphere (i.e., zenith, north celestial pole, celestial equator).
Elaborate	What other activity(ies) can my students do that will allow them to explore this activity in a variety of contexts?	After helping students become comfortable with views of the apparent motions of stars from latitudes $0^\circ$ and $90^\circ$ , the instructor proposed views from the latitude of the home university as well as a southern latitude. Students were again asked to observe, discuss, and record characteristics seen from each latitude presented.
Evaluate	How can I evaluate if my students have reached the intended learning objectives?	Students' learning was evaluated through their ability to respond to 2–4 assessments.

**APPENDIX B**

In Appendix B, the learning objectives of each lab in the introductory astronomy laboratory described in this study are presented.

TABLE XI. Learning objectives of each lab in the introductory astronomy laboratory described in this study.

Lab number and title	Learning objectives
Lab 1: Characterizing Earth and Sky	<ol style="list-style-type: none"> <li>1. Uses angles to specify locations on spheres.</li> <li>2. Describes the role of the Earth in limiting a person’s view of the celestial sphere.</li> <li>3. (skill) Uses the four quantities (altitude, compass direction, magnitude, and angles between stars) to describe a constellation’s position and appearance.</li> <li>4. (skill) Determines the direction North from a randomly oriented SF horizon, given a clear night sky.</li> </ol>
Lab 2: Watching Stars from Earth	<ol style="list-style-type: none"> <li>1. Determines observer’s latitude using Polaris.</li> <li>2. Given the position of Polaris, predicts which stars rise and set.</li> <li>3. Predicts how stars move with respect to Polaris over the course of 24 hours.</li> <li>4. Explains how the visible portion of the celestial sphere changes depending on latitude.</li> </ol>
Lab 3: Sunrise Sunset	<ol style="list-style-type: none"> <li>1. Describes Sun’s path across the sky over the course of one day.</li> <li>2. Predicts variations in the Sun’s rise and set locations over the course of a year.</li> <li>3. Predicts changes in the Sun’s maximum altitude over the course of a year.</li> <li>4. Given a date, e.g., a solstice or equinox and a latitude, deduces the Sun’s altitude at noon.</li> <li>5. Explains why the Sun’s rise and set location depends on Earth’s location in its orbit and one’s latitude.</li> </ol>
Lab 4: Changes in the Night Sky throughout the Year	<ol style="list-style-type: none"> <li>1. Conceptualizes the role of the Earth’s orbit in creating changes in view.</li> <li>2. Predicts daily and monthly changes of the Sun’s location along the ecliptic using the Zodiacal constellations as a reference.</li> <li>3. Predicts daily and monthly changes in the stars rise and set times.</li> </ol>
Lab 5: Phases of the Moon	<ol style="list-style-type: none"> <li>1. Explains why we see the Moon go through phases.</li> <li>2. Given a Moon phase, predicts the phase of the Moon a given number of days later or earlier.</li> <li>3. Given a Moon phase, predicts the approximate rise and set time, or vice versa.</li> </ol>

**APPENDIX C**

In Appendix C, the assessment measuring students’ ability to predict the number of stars that go below the horizon when given the position of Polaris is presented. The rubric used to analyze students’ responses to the open-ended question on this assessment is also presented.

You look north out your window and see the sky as drawn in the diagram. How many of the stars shown will be below the horizon sometime over the next 24 hours?

←————— Horizon —————→

Explain your reasoning....

- A. None
- B. 1
- C. 2
- D. 3
- E. 4
- F. 5
- G. All will be below the horizon sometime over 24 hours

FIG. 9. Assessment used to measure students’ ability to predict the number of stars that go below the horizon when given the position of Polaris. The distributions of students choosing the correct multiple-choice response and attaining the minimum rubric score can be found in Fig. 2.

TABLE XII. Rubric used to determine rubric scores for open-ended responses given to the assessment shown in Fig. 9.

Rubric categories (Minimum Rubric Score: 2/4)	0	1
A. Draws or mentions that stars travel in circles around Polaris	No	Yes
B. Draws or explains that the three outermost stars will be the stars to go under the horizon because they have a greater separation from the horizon than Polaris	No	Yes
C. Describes the stars motion as apparent	No	Yes
D. Mentions Earth's rotation as the cause of stars motion	No	Yes

### APPENDIX D

In Appendix D, the assessment measuring students' ability to identify how the direction and length of their shadow change over a given time period is presented. The rubric used to analyze students' responses to the open-ended question on this assessment is also presented.

You observe your shadow at noon in the continental United States on the first day of October.  
Two weeks later your shadow will appear...

- Shorter and will point toward the south.
- Shorter and will point toward the north.
- Longer and will point toward the south.
- Longer and will point toward the north.

Explain your reasoning...

FIG. 10. Assessment used to measure students' ability to identify how the direction and length of their shadow change over a given time period. The distributions of students choosing the correct multiple-choice response and attaining the minimum rubric score can be found in Fig. 3. This question was adapted from the Center for Astronomy Education exam bank.

TABLE XIII. Rubric used to determine rubric scores for open-ended responses given to the assessment shown in Fig. 10.

Rubric categories (Minimum Rubric Score: 4/6)	0	1	2	3
A. North or South reasoning	Does not mention the Sun as a contributor to a shadow	Sun causes shadow	Sun's position is in the South at noon OR Your shadow will point in the opposite direction of the Sun	Sun's southerly position causes a shadow that points north
B. Longer or shorter reasoning	Does not mention the Sun as a contributor to the length of shadow	The height of the Sun in the sky determines the length of shadow	The higher the Sun is in the sky the shorter your shadow will be OR The lower the Sun is in the sky the longer your shadow will be	Two weeks later from the time identified at noon, the Sun will be lower in the sky which causes a longer shadow

**APPENDIX E**

In Appendix E, the assessment measuring students’ ability to identify the portion of the celestial sphere that is visible from various locations on Earth is presented. The rubric used to analyze students’ responses to the open-ended question on this assessment is also presented.

To see the greatest number of stars possible throughout the period of one year, an ideal location for a person to live would be under dark skies....

- A. At the North Pole (90° N).
- B. At a latitude similar to San Francisco (38° N)
- C. Near the equator (0° N)
- D. All locations on Earth see the same number of stars.

Explain your reasoning...

FIG. 11. Assessment used to measure students’ ability to identify the portion of the celestial sphere that is visible from various locations on Earth. The distributions of students choosing the correct multiple-choice response and attaining the minimum rubric score can be found in Fig. 4. This question was adapted from Paul Green’s *Peer Instruction for Astronomy*.

TABLE XIV. Rubric used to determine rubric scores for open-ended responses given to the assessment shown in Fig. 11.

Rubric categories (Minimum Rubric Score: 2/3)	0	1
A. Describes or draws that at the equator...you can see all stars OR you can see stars from both the Northern and Southern hemisphere of the Celestial sphere	No	Yes
B. Describes or draws that at another location on Earth besides the Equator.. you see fewer stars OR you see a smaller portion of the Celestial Sphere.	No	Yes
C. Describes that your position on Earth relative to the Earth’s rotational axis determines the number of stars you see over the course of a year	No	Yes

**APPENDIX F**

In Appendix F, the assessment measuring students’ ability to provide a description of how and why the night sky changes over the course of six months is presented. The rubric used to analyze students’ responses to the open-ended question on this assessment is also presented.

**Circle** a number below that reflects your agreement/disagreement with this statement:

“The pattern of stars seen in the sky from San Francisco tonight at 11PM will be the same pattern of stars seen at 11 PM in six months from now.”

**Disagree      Agree      Don’t Know**

**Please explain your thoughts regarding your choice.  
In your response, describe in detail what about this statement you agree or disagree with.**

USE A DRAWING to help aid your explanation.

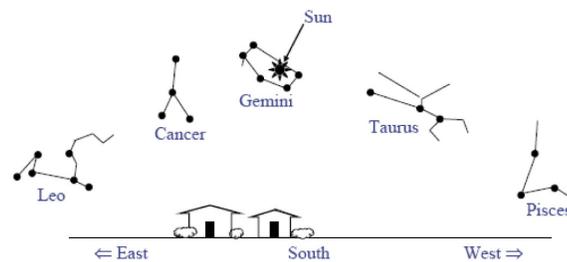
FIG. 12. Assessment used to measure students’ ability to provide a description of how and why the night sky changes over the course of six months. The distributions of students choosing the correct multiple-choice response and attaining the minimum rubric score can be found in Fig. 5. This question was adapted from Paul Green’s *Peer Instruction for Astronomy*.

TABLE XV. Rubric used to determine rubric scores for open-ended responses given to the assessment shown in Fig. 12.

Rubric categories (Minimum Rubric Score: 3/4)	0	1
A. Mentions that the sky or the view of the night sky will look different six months from now	No	Yes
B. Mentions Earth's motion around the Sun or draws Earth's orbit	No	Yes
C. Describes that the position of Earth on its orbit around the Sun determines an observer's view of the night sky	No	Yes
D. Mentions or draws that the direction you face at night six months from now is different	No	Yes

## APPENDIX G

In Appendix G, the assessment measuring students' ability to describe the Sun's daily motion versus annual motion is presented. The rubric used to analyze students' responses to the open-ended question on this assessment is also presented.



Use the figure above to answer this question.

If you could see the stars during the day, the drawing above shows what the sky would look like at noon on a given day. The Sun is in front of the stars of the constellation Gemini. Which constellation would you expect the Sun to be located in front of when it sets that day?

- A. Leo
- B. Cancer
- C. Gemini
- D. Taurus
- E. Pisces

Explain your reasoning...

FIG. 13. Assessment used to measure students' ability to describe the Sun's daily motion versus annual motion. The distributions of students choosing the correct multiple-choice response and attaining the minimum rubric score can be found in Fig. 6. This question was adapted from the Astronomy Diagnostic Test.

TABLE XVI. Rubric used to determine rubric scores for open-ended responses given to the assessment shown in Fig. 13.

Rubric categories (Minimum Rubric Score: 4/8)	0	1	2
A. Earth’s rotation in creating apparent motion of Sun (mechanistic)	Does not address rotation as a contributor to position of Sun in stars	Mentions Earth’s rotation in their answer	Describes Earth’s rotation as the reason for the Sun’s and/or stars motion over the course of a day
B. Earth’s orbit in creating apparent motion of Sun (mechanistic)	Does not mention Earth’s orbit as a contributor to appearance of Sun with regards to the stars	Mentions or draws Earth’s orbit in their answer	Describes Earth traveling on its orbit as the motion that causes the Sun to appear in front of different constellations
C. Appearance of the Sun and stars over the course of a day (descriptive)	Does not address relationship between position of Sun and stars	Mentions that the Sun sets in the west	Describes that the Sun and stars travel together
		OR	OR
		Mentions that the Sun rises in the east	Describes that the Sun will be in front of Gemini (or the same group of stars) when it sets that day
		OR	OR
		Mentions that Gemini will set in the west	Describes that the Sun will stay in front of the same group of stars over the course of a day
D. Appearance of the Sun and stars over periods longer than a day (descriptive)	Does not address relationship between position of Sun and stars	Mentions that the Sun will appear in front of the same constellation over the course of a month	Describes that the Sun will spend about a month in front of each of the constellations along the ecliptic or zodiac

**APPENDIX H**

In Appendix H, the assessment measuring students’ understanding of Moon phases as a result of Earth’s perspective of the Moon’s half-lit side is presented. The rubric used to analyze students’ responses to the open-ended question on this assessment is also presented.

**Circle** a number below that reflects your agreement/disagreement with this statement:

*“The reason the Moon has phases is because sometimes Earth blocks the Sun’s light from reaching the Moon.”*

**Disagree                      Agree    Don’t Know**

**Please explain your thoughts regarding your choice.  
In your response, describe in detail what about this statement you agree or disagree with.**

FIG. 14. Assessment used to measure students’ understanding of Moon phases as a result of Earth’s perspective of the Moon’s half-lit side. The distributions of students choosing the correct multiple-choice response and attaining the minimum rubric score can be found in Fig. 7.

TABLE XVII. Rubric used to determine rubric scores for open-ended responses given to the assessment shown in Fig. 14.

Rubric categories (Minimum Rubric Score: 2/4)	0	1
A. Draws or mentions that the orbital plane of the Moon is tilted with respect to the Earth-Sun plane	No	Yes
B. Mentions that the Moon is always half lit by the Sun	No	Yes
C. Describes that it is Earth’s vantage point of the half lit portion of the Moon that creates the appearance of phases	No	Yes
D. Describes an eclipse (or a lunar eclipse) as the only time when the Earth blocks the light of the Sun from hitting or reaching the Moon	No	Yes

## APPENDIX I

In Appendix I, the assessment measuring students' ability to predict the phase of the Moon in two weeks given the current time and its current position on the observer's horizon is presented. The rubric used to analyze students' responses to the open-ended question on this assessment is also presented.

The Moon is on the Earth's eastern horizon at 3AM tonight. In two weeks what will the phase of the Moon be?

- A. Waning crescent
- B. Waxing gibbous
- C. Waning gibbous
- D. Third quarter
- E. New

Explain your reasoning. Use a drawing to aid your explanation....

FIG. 15. Assessment used to measure students' ability to predict the phase of the Moon in two weeks given the current time and its current position on the observer's horizon. The distributions of students choosing the correct multiple-choice response and attaining the minimum rubric score can be found in Fig. 8. This question was adapted from the Center for Astronomy Education exam bank.

TABLE XVIII. Rubric used to determine rubric scores for open-ended responses given to the assessment shown in Fig. 15.

Rubric categories (Minimum Rubric Score: 2/4)	0	1
A. Determines the phase of the Moon at 3AM tonight (waning crescent)	No	Yes
B. Includes a drawing that accurately depicts that the phase of the Moon will be a waning crescent tonight at 3AM	No	Yes
C. Determines how far the Moon travels along its orbit in two weeks	No	Yes
D. From the phase given for the current time, writes what the phase of the Moon will be in two weeks	No	Yes

- [1] B. Bridgeman and R. Morgan, Success in college for students with discrepancies between performance on multiple-choice and essay tests, *J. Educ. Psychol.* **88**, 333 (1996).
- [2] M. Anbar, Comparing assessments of students' knowledge by computerized open-ended and multiple-choice tests, *Acad. Med.* **66**, 420 (1991).
- [3] H. K. Rabinowitz and M. Hojat, A Comparison of the modified essay question and multiple choice question formats: Their relationship to clinical performance, *Fam. Med.* **21**, 364 (1989).
- [4] D. W. Hudgins, E. Prather, D. Grayson, and D. Smits, Effectiveness of collaborative Ranking Tasks on student understanding of key astronomy concepts, *Astron. Educ. Rev.* **5**, 1 (2006).
- [5] C. S. Wallace, E. Prather, and D. Duncan, A study of general education astronomy students' understanding of cosmology. Part V. The effects of a new suite of cosmology Lecture-Tutorials on students' conceptual knowledge, *Int. J. Sci. Educ.* **34**, 1297 (2012).
- [6] H.-S. Lee, O. L. Liu, and M. C. Linn, Validating measurement of knowledge integration in science using multiple-choice and explanations items, *Appl. Measur. Edu.* **24**, 115 (2011).
- [7] R. W. Bybee, J. Taylor, A. Gardner, P. Van Scotter, J. Powell, A. Westbrook, and N. Landes, Biological Sciences Curriculum Study on The BSCS 5E Instructional Model: Origins and Effectiveness, 2006 (unpublished).
- [8] G. J. Posner, K. A. Strike, P. W. Hewson, and W. A. Gertzog, Accommodation of a scientific conception: Toward a theory of conceptual change, *Sci. Educ.* **66**, 211 (1982).
- [9] G. P. Wiggins and J. McTighe, *Understanding by Design*, 2nd ed. (Association for Supervision and Curriculum Development, Alexandria, VA, 2005).
- [10] B. Hufnagel, Development of the Astronomy Diagnostic Test, *Astron. Educ. Rev.* **1**, 47 (2002).
- [11] P. J. Green, *Peer Instruction for Astronomy* (Addison Wesley, Boston, MA, 2002).

- [12] F. Faul, E. Erdfelder, A. Lang, and A. Buchner, G\* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences, *Behav. Res. Meth. Instrum. Comput.* **39**, 175 (2007).
- [13] J. M. Hoenig and D. Heisey, The abuse of power: The pervasive fallacy of power calculations for data analysis, *Am. Stat.*, **55**, 19 (2001).
- [14] E. M. Barder, E. Prather, K. Brecher, and T. Slater, Development and Validation of the Light and Spectroscopy Concept Inventory, *Astron. Educ. Rev.* **5**, 103 (2006).
- [15] J. M. Bailey and T. Slater, Finding the forest amid the trees: Tools for evaluating astronomy education and public outreach projects, *Astron. Educ. Rev.* **3**, 47 (2004).