A LIVELY ELECTRONIC COMPENDIUM OF RESEARCH, NEWS, RESOURCES, AND OPINION

Astronomy Education Review

2011, AER, 10, 010106-1, 10.3847/AER2011029

A Study of General Education Astronomy Students' Understandings of Cosmology. Part I. Development and Validation of Four Conceptual Cosmology Surveys

Colin S. Wallace
Center for Astronomy Education (CAE), Steward Observatory, University of Arizona, Tucson, Arizona 85721
Edward E. Prather
Center for Astronomy Education (CAE), Steward Observatory, University of Arizona, Tucson, Arizona 85721
Douglas K. Duncan
Department of Astrophysical and Planetary Sciences, University of Colorado at Boulder, Boulder, Colorado 80309
Received: 09/21/11, Accepted: 11/1/11, Published: 12/15/11

Abstract

This is the first in a series of five articles describing a national study of general education astronomy students' conceptual and reasoning difficulties with cosmology. In this paper, we describe the process by which we designed four new surveys to assess general education astronomy students' conceptual cosmology knowledge. These surveys focused on the expansion and evolution of the universe, the Big Bang, and the evidence for dark matter in spiral galaxies. We also present qualitative evidence for the validity of these surveys.

1. INTRODUCTION

The general education introductory astronomy course (hereafter Astro 101) is the final science course many of our future teachers, business leaders, journalists, and politicians ever take (Deming and Hufnagel 2001; Rudolph *et al*. 2010). While much of astronomy education research has focused on improving student learning at the Astro 101 level, many gaps exist in our knowledge of students' conceptual and reasoning difficulties relating to key topics in astronomy. Despite the fact that cosmology is one of the most commonly taught topics in all of Astro 101 (Slater *et al*. 2001), we have, with the exception of a handful of studies (Comins 2001; Hansson and Redfors 2006; Lightman and Miller 1989; Lightman Miller, and Leadbeater 1987; Prather, Slater, and Offerdahl 2002; Simonelli and Pilachowski 2003), very little information on what Astro 101 students struggle with when they study cosmology.

This is the first in a series of five papers describing one of the first large-scale, systematic efforts to uncover Astro 101 students' learning difficulties with cosmology (but see also Bailey *et al*. 2011 and Coble *et al*. 2011 for results from a complementary study). This work concentrates on answering two questions:

(1) What are the common conceptual and reasoning difficulties Astro 101 students encounter when they study cosmology?
(2) Do students who use the new suite of cosmology *Lecture-Tutorials* (which we also designed, piloted, and validated as part of this study) achieve larger learning gains than their peers who did not?

This five paper series explicates our study and addresses these two research questions. This first paper focuses on the process by which we designed and validated four surveys used to measure Astro 101 students' conceptual knowledge of commonly taught cosmology topics. Subsequent papers quantitatively evaluate these surveys from the perspectives of classical test theory (Wallace, Prather, and Duncan 2011a, aka "Paper 2") and item response theory (Wallace, Prather, and Duncan 2011b, aka "Paper 3"). Wallace, Prather, and Duncan (2011c; "Paper 4") report students' most common difficulties with cosmology, as uncovered by our study. Wallace, Prather, and Duncan (2011d; "Paper 5") describe how a *Lecture-Tutorial* approach to teaching cosmology can lead to significantly increased student understanding over traditional instruction.

This paper is organized as follows. Section 2 describes which cosmology topics we decided to focus on and why. Section 3 presents the process by which we constructed the surveys. We outline the four conceptual cosmology surveys in Section 4. In Section 5, we present evidence for the validity of these surveys. Section 6 summarizes the results of this portion of our study.

2. TOPIC SELECTION

While several surveys of Astro 101 students' conceptual knowledge already exist (e.g., Bailey 2007; Bailey, Slater, and Slater 2011; Bardar *et al*. 2007; Hufnagel 2002; Keller 2006; Lindell 2001; Sadler *et al*. 2010; Trumper 2000), none of them focus exclusively on cosmology. Thus, one of our first steps was to create a suite of questions that can be used to probe students' understandings of central ideas in cosmology. Because modern cosmology is such a broad field, we necessarily restricted the scope of our questions to a subset of cosmological topics. Specifically, we are interested in topics that share two or more of the following characteristics:

(1) They are frequently taught in Astro 101 (see Slater, Adams, Brissenden, and Duncan 2001 as well as introductory-level textbooks such as Bennett *et al*. 2008, Chaisson and McMillan 2011, and Seeds and Backman 2011).
(2) They are conceptually complex, yet accessible to Astro 101 students (for example, we think that Astro 101 students should be able to discuss the expansion of the universe and the Big Bang at the level of Lineweaver and Davis 2005).
(3) Previous studies indicate the potential for widespread difficulties and confusion related to the topic (e.g. students' incorrect ideas about the Big Bang in Prather, Slater, and Offerdahl 2002).

Based on these criteria, we decided to focus on the following topics: the expansion and evolution of the universe, the Big Bang, and the evidence for dark matter in spiral galaxies.

With regards to the third criteria, several previous studies imply that these topics might be especially troublesome for Astro 101 students. For example, the closely related topics of the expansion of the universe and the Big Bang appear to be widely misunderstood or even unknown. Lightman, Miller, and Leadbeater (1987) and Lightman and Miller (1989) report that most people are unaware that the universe is expanding and, in the absence of any evidence to the contrary, presume it is static. Lightman, Miller, and Leadbeater (1989) further report that 43% of the high school students they surveyed expressed fear or other negative psychological associations with the prospect of an expanding universe. Prather, Slater, and Offerdahl (2002), Hansson and Redfors (2006), and Lineweaver and Davis (2005) describe how people think of the Big Bang as an explosion of pre-existing matter into empty space, as opposed to the beginning of the expansion of space. (Note that while Lineweaver and Davis 2005 provide an excellent discussion of the relevant physics, they did not conduct any research on people's naïve ideas about the Big Bang, unlike Prather, Slater, and Offerdahl 2002 and Hansson and Redfors 2006.) Prather, Slater, and Offerdahl (2002) hypothesize that students' belief in pre-existing matter may be a manifestation or activation of a phenomenological primitive (diSessa 1993) which they call "you can't make something from nothing." Prather, Slater, and Offerdahl (2002) as well as Simonelli and Pilachowski (2003) also note that a sizeable minority of students associate Earth, the Solar System, and the formation of the Solar System with their understandings of the Big Bang. Taken together, these papers suggest multiple issues with Astro 101 students' ideas about the expansion of the universe and the Big Bang.

Previous studies also suggest that Astro 101 students may have trouble reading and interpreting both Hubble plots and galaxy rotation curves. Hubble plots graph galaxy recession velocities as a function of distance and are used to infer the age and expansion rate of the universe. Astronomers infer the presence of dark energy due to the changing expansion rate represented by the changing slope in our universe's Hubble plot. A galaxy's rotation curve plots the orbital velocities of objects in that spiral galaxy as a function of distance from its galactic center. Astronomers infer that dark matter dominates galaxies, in part, because galaxy rotation curves are flat at large radii. Astro 101 students are frequently exposed to both Hubble plots and galaxy rotation curves, and both require students to understand kinematic and dynamic quantities, such as velocity and force, and how to infer other quantities through the analysis of graphs. Trowbridge and McDermott (1980, 1981) elucidate several reoccurring difficulties students experience with kinematic quantities, and McDermott, Rosenquist, and van Zee (1987) highlight multiple errors students frequently make when reading graphs. These prior studies imply that Astro 101 students may struggle to correctly read and interpret Hubble plots and galaxy rotation curves.

In some cases, we are interested in students' ideas on cosmological topics that have not been addressed by any prior research. These include the evolution of the universe's density and temperature with time, as well as expansion's effect on the relationship between lookback times and distances. With respect to this last point, Davis and Lineweaver (2004) and Lineweaver and Davis (2005) explain (in more and less technical detail, respectively) how the farthest object we can see is currently about 46 billion light-years away from us (and receding many times faster than the speed of light) even though the universe is only about 13.7 billion years old (Spergel *et al*. 2007). While conceptual treatments of these issues are found in introductory-level textbooks (e.g. Bennett *et al*. 2008) as well as articles written for the general public (Lineweaver and Davis 2005), there is little research on what learning experiences students need to go through in order to understand these non-intuitive ideas.

3. SURVEY DESIGN

To design our conceptual cosmology surveys, we followed the four-step process of survey design and interpretation recommended by Wilson (2005):

1) Define the constructs to be measured and create construct maps for each construct;
2) Design survey items;
3) Categorize and score the full range of responses; and
4) Apply psychometric models to the data.

This paper primarily focuses on the first two steps, while subsequent papers in this series focus on the last two.

3.1. Constructs and Construct Maps

The term *construct* refers to "the concept or characteristic that a test is designed to measure" (AERA, APA, and NCME 1999, p. 5; see also Wilson 2005). There is widespread agreement that developing a robust and specific definition of a construct is always the first step one should take in survey development (Gorin 2006; Shepard 1993; Wilson 2005). By defining the construct of interest in advance, survey designers can select items that accurately probe multiple levels and/or multiple components of the construct (Gorin 2006; Shepard 1993; Wilson 2005). The key to good survey design is to operationalize, to the greatest extent possible, the construct the survey is supposed to measure.

One way to operationalize what a survey measures is to create a *construct map* for each survey (Briggs *et al*. 2006; Wilson 2005). A construct map is a table that shows the different levels into which students may fall on the construct, since some students will have more expert-like understandings of the relevant construct than others. The construct map is a visual or tabular instantiation of a construct and how people may vary on that construct (Wilson 2005). Note that a construct map does not necessarily imply a learning progression – students do not necessarily pass through all levels as they develop more expert-like understandings. For examples of construct maps, see Briggs *et al*. (2006), Wilson (2005), and Tables 1–4 below.

Why does this process of construct definition and construct map creation matter? Because the validity of a survey – that is, whether or not the survey actually measures what its designers think it measures – depends on the process by which one selects items for inclusion or exclusion from the survey (Briggs *et al*. 2006). Part of the validity argument for a survey depends on establishing whether that survey's items adequately cover the survey's construct (Gorin 2006; Shepard 1993). By creating construct maps for each construct, we can ensure that we develop surveys that can place students at all levels on the construct map. Thus, the process underlying the construction of the survey is a necessary (although not sufficient – see, for example, Shepard 1993) condition for establishing the validity of the survey.

As Section 2 implies, this study encompasses multiple constructs. Specifically, we created four surveys, each of which focuses on one of the following four constructs of interest:

Table 1. Construct map for the *Hubble's plots* construct

Level	Description
3	The student ***correctly*** reasons about the age and the expansion rate of the universe using Hubble plots.
2	The student ***sometimes correctly*** reasons about Hubble plots, but ***sometimes cues on the wrong features of the graph or incorrectly interprets a feature***.
1	The student ***incorrectly*** reasons about the age and the expansion rate of the universe using Hubble plots.

1) Interpreting Hubble plots;
2) Models of the expansion of the universe and the Big Bang;
3) The evolving universe; and
4) Evidence for dark matter in spiral galaxies.

The abbreviated names *Hubble plots*, *models*, *evolving universe*, and *dark matter*, respectively, hereafter refer to these constructs. They are defined and described in more detail below, along with their associated construct maps. All of the construct maps shown here represented our hypotheses about the different levels of mastery into which students may fall on their associated constructs.

The *Hubble plots* construct looks at whether or not a student can use Hubble plots to reason about the age and expansion rate of the universe. Since this construct strongly relies on students' abilities to interpret graphs, our work was informed by the work of McDermott, Rosenquist, and van Zee (1987) on student difficulties in reading graphs. McDermott, Rosenquist, and van Zee (1987) found that students often focus on an inappropriate feature of a graph (for example, they might look at the height of a graph when they need to focus on its slope and vice-versa). Students also struggle to interpret negative quantities, such as velocity and acceleration. Finally, they found that students tend to think that the graph of an object's motion should resemble the path of its motion (e.g. a ball rolling down a hill should have a velocity versus time graph that is shaped like the hill). We hypothesized that Astro 101 students encounter similar difficulties reading Hubble plots, especially because Hubble plots include information on kinematic quantities such as velocity and acceleration, which present their own sets of difficulties for students (Trowbridge and McDermott 1980, 1981).

Table 2. Construct map for the *models* construct

Level	Description
4	The student ***correctly*** states that the universe is physically expanding over time. The student ***correctly*** states that only galaxies are moving apart from one another due to expansion. The student ***correctly*** claims that the universe has no center. The student ***correctly*** claims that the universe has no edge. The student ***correctly*** describes the Big Bang as the beginning of expansion.
3	The student ***correctly*** states that the universe is physically expanding over time. The student ***incorrectly*** states that all objects in the universe are moving apart from one another due to expansion. The student ***correctly*** claims that the universe has no center. The student ***correctly*** claims that the universe has no edge. The student ***may*** describe the Big Bang as an explosion ***or*** as the beginning of expansion.
2	The student ***correctly*** states that the universe is physically expanding over time. The student ***incorrectly*** states that all objects in the universe are moving apart from one another due to expansion. The student ***incorrectly*** claims that the universe has a center. The student ***incorrectly*** claims that the universe has an edge. The student ***incorrectly*** describes the Big Bang as an explosion but ***not*** as the beginning of something smaller than the universe (e.g. the Solar System, Galaxy, etc.).
1	The student ***incorrectly*** states that the universe is not physically expanding over time. The student ***may or may not*** claim that the universe has a center. The student ***may or may not*** claim that the universe has an edge. The student ***incorrectly*** describes the Big Bang as an explosion and/or as the beginning of something smaller than the universe (e.g. the Solar System, Galaxy, etc.).

Table 3. Construct map for the *evolving universe* construct

Level	Description
3	The student ***correctly*** relates the light travel time between two galaxies to their past, present, or future distances from one another. The student ***correctly*** describes how the temperature of the universe has changed over time. The student ***correctly*** describes how the density of matter in the universe has changed over time. The student ***correctly*** states that the matter in the universe has ***not*** always existed.
2	The student ***correctly*** relates the light travel time between two galaxies to their past, present, or future distances from one another. The student ***correctly*** describes how the temperature of the universe has changed over time. The student ***correctly*** describes how the density of matter in the universe has changed over time. The student ***incorrectly*** states that the matter in the universe has always existed.
1	The student ***incorrectly*** relates the light travel time between two galaxies to their past, present, or future distances from one another. The student ***incorrectly*** describes how the temperature of the universe has changed over time. The student ***incorrectly*** describes how the density of matter in the universe has changed over time. The student ***incorrectly*** states that the matter in the universe has always existed.

Table 1 is the construct map for the *Hubble plots* construct. At the lowest level (Level 1) are students that always incorrectly interpret Hubble plots. At the highest level (Level 3) are students that correctly use Hubble plots to qualitatively reason about the expansion rate and age of the universe. Students who sometimes correctly reason using Hubble plots fall in the middle region (Level 2). We hypothesized that students who fall into Level 1 or Level 2 make many of the graph interpretation errors outlined by McDermott, Rosenquist, and van Zee (1987). Additionally, we hypothesize that students in Levels 1 and 2 may also neglect that the farther away one looks in the universe, the further back in time one sees. Level 1 or 2 categories would indicate a learner who does not read large distances on a Hubble plot as times far in the past to arrive at an understanding of when events (e.g. changes in the expansion rate) occur throughout the history of the universe. This requirement adds an

Table 4. Construct map for the *dark matter* construct

Level	Description
4	The student ***correctly*** identifies the galaxy rotation curve of a spiral galaxy. The student ***correctly*** describes how the orbital speeds of stars at different radii relate to one another based on the galaxy rotation curve s/he chose. The student ***correctly*** describes how mass is distributed in the galaxy based on the galaxy rotation curve s/he chose.
3	The student ***incorrectly*** identifies the galaxy rotation curve of a spiral galaxy. The student ***correctly*** describes how the orbital speeds of stars at different radii relate to one another based on the galaxy rotation curve s/he chose. The student ***correctly*** describes how mass is distributed in the galaxy based on the galaxy rotation curve s/he chose.
2	The student ***incorrectly*** identifies the galaxy rotation curve of a spiral galaxy. The student ***correctly*** describes how the orbital speeds of stars at different radii relate to one another based on the galaxy rotation curve s/he chose. The student ***incorrectly*** describes how mass is distributed in the galaxy based on the galaxy rotation curve s/he chose.
1	The student ***incorrectly*** identifies the galaxy rotation curve of a spiral galaxy. The student ***incorrectly*** describes how the orbital speeds of stars at different radii relate to one another based on the galaxy rotation curve s/he chose. The student ***incorrectly*** describes how mass is distributed in the galaxy based on the galaxy rotation curve s/he chose.

astronomical twist to the difficulties discussed in McDermott, Rosenquist, and van Zee (1987). In Section 3.2, we describe the process by which we designed items that help us place students at their appropriate levels on the construct map.

The *models* construct focuses on students' conceptualizations of the expansion of the universe and the Big Bang. Much of the previous research on conceptual difficulties with cosmology applies to this construct (Comins 2001; Hansson and Redfors 2006; Lightman, Miller, and Leadbeater 1987; Lightman and Miller 1989; Prather, Slater, and Offerdahl 2002; Simonelli and Pilachowski 2003). These previous studies influenced our design of the construct map for this construct.

The *models* construct map is shown in Table 2. People who do not know that the universe is expanding, or that the Big Bang is related to the expansion of the universe, are at the lowest level (Level 1). At Level 2 are people who conceive of expansion and the Big Bang as the motion of objects in the universe away from a center into empty space. At the highest levels (Levels 3 and 4) are those who relate the expansion of the universe and the Big Bang to the expansion of space itself. People at Level 3 are almost identical to people at Level 4 with two exceptions. One exception is that those at Level 3 erroneously claim that *all* objects in the universe – planets, starts, galaxies, etc. – move away from one another due to the expansion of the universe. In general, only the distances between widely separated galaxies are affected by the expansion of the universe. Although this construct map is based, in part, on prior studies, we must emphasize that it, like all the construct maps originally presented here, originally represented a hypothesis about how students are arranged along this construct.

The *evolving universe* construct looks at whether or not a student knows how properties of the universe have changed over time. Specifically, this construct focuses on students' knowledge of how expansion has affected the amount of matter in the universe, the density of matter in the universe, the temperature of the universe, and the relationship between lookback times, proper distances, and light travel times between widely separated galaxies. The *evolving universe* construct map is shown in Table 3.

The *dark matter* construct probes whether a student can construct the causal chain of reasoning linking the flat galaxy rotation curves of spiral galaxies to the existence of dark matter. The construct map for this construct is displayed in Table 4.

The construct map in Table 4 has four levels. At the lowest level (Level 1) are people who do not demonstrate any correct link in the chain of reasoning mentioned above. At Level 2 are people who, despite selecting an incorrect galaxy rotation curve, correctly relate the orbital speeds of stars at various radii using that galaxy rotation curve. If someone can also connect the orbital speeds of stars to the distribution of mass in the galaxy, then the student will be at Level 3. Level 4 is reserved for students who select the right galaxy rotation curve and correctly relate it to the orbital speeds of stars and the distribution of mass in the galaxy. We hypothesized that some students will pick the wrong galaxy rotation curve, but then correctly use that galaxy rotation curve to connect the orbital speeds of stars and the distribution of mass in the galaxy. Since connecting these three ideas is likely nontrivial for the average Astro 101 students, a student who can make such a connection should be placed high on the construct map. After all, the student may understand the relevant physics but not know what the true galaxy rotation curve of a spiral galaxy looks like, much like astronomers several decades ago.

Informed by these constructs and construct maps, we designed items that help us to place students at their appropriate levels on the construct maps. We describe the principles underlying the design of our items in Section 3.2.

3.2. Item Design

In some respects, the process of designing items involves a certain amount of creativity or inspiration on the part of the survey designer (Wilson 2005). However, such "item brainstorming sessions" must be complemented by a detailed evaluation of whether or not the items adequately cover the construct of interest (Shepard 1993). Furthermore, there are potentially an infinite (or at least a very large) number of items one could include in the survey. Wilson (2005) describes the process of item design as selecting items from the pool of potential items. He goes on to define two key components of items that help determine the item pool: 1) the construct component and 2) the descriptive component.

The construct component of a survey's items reflects the degree to which the items help place respondents at various levels of the construct (Wilson 2005). When designing an item, one should always ask "If a student answers this item correctly or incorrectly, what does that tell me about the amount of the construct she 'possesses'?" The construct map associated with a construct should specify the attributes a survey's items need to elicit from respondents (Wilson 2005).

The descriptive component refers to all the other required properties of the items that do not relate to the construct (Wilson 2005). For example, should the survey include only free-response items, only multiple-choice items, or a combination of both? How quickly can the average student provide a complete and correct response? Is the use of jargon essential or superfuous? One must answer these and related questions when reducing the item pool.

For the four surveys we wrote, we were cognizant of several constraints to the descriptive component of our items. First, the majority of items have to be open-ended. The responses to open-ended questions have the potential to reveal common reasoning difficulties held by students, as well as reveal students' natural language; this is why many multiple-choice survey instruments begin as open-ended questions early in their development phase (e.g. Bailey 2007; Bardar et al. 2007). Given the lack of research on what Astro 101 students struggle with regarding topics in cosmology, we felt we needed more information before we could create effective multiple-choice items. However, asking too many open-ended questions quickly leads to student fatigue and a corresponding degradation in item responses. Additionally, we found that instructors were generally only willing to give up about twenty minutes of class time for a survey. One solution is to provide students with response options (thereby limiting the amount of time they need to indicate their answers) while still asking them to explain why they chose their answer (providing the benefit of an open-ended question). This is the approach we took with a subset of items. Furthermore, we limited the number of items on each survey. In general, each survey contained one item per attribute on its associated construct map. While more items would undoubtedly increase the accuracy and precision with which we place a student at a given level on a construct map, they would also lengthen the test and reduce the efficacy of individual questions.

Our survey items are subject to other constraints as well. In order to use the surveys for pre-post testing, the questions have to be worded using language that is accessible to students regardless of whether or not they have received instruction. Yet the question must still somehow probe a student's understanding. We here adopt Heron's (2004) definition of understanding: A student understands a topic "if, when faced with an unfamiliar problem, he or she reliably selects the appropriate concepts and principles, applies them correctly, and constructs a logically sound solution" (p. 342). We could not restrict our item pool to items that simply elicit factual recall since students may correctly answer such items without understanding the underlying concepts (Vosniadou 1994). This suggested creating questions modeled after what Vosniadou (1994) calls generative questions. A generative question confronts people "with phenomenon about which they do not have any direct experience and about which they have not yet received explicit instruction. Because generative questions cannot be answered through the simple repetition of unassimilated information, they have a greater potential for unraveling underlying conceptual structures" (Vosniadou 1994, p. 50). Where possible, we used generative questions written with non-technical language (although some astrophysical terms and jargon were unavoidable). We also used several well-established techniques of question writing. Each item (when appropriate) probes a single idea, avoids negatives, provides space between multiple parts of a single question, and bolds, italicizes, and/or underlines important words (Henriques, Colburn, and Ritz 2006). These techniques were used in order to clarify what each item is asking.

4. OVERVIEW OF SURVEYS

As noted above, we ultimately wrote four surveys. Each survey is designed to measure Astro 101 students' conceptual knowledge for one of the four constructs described above. All of our surveys went through multiple cycles of testing and revision over the course of this study. We describe some of the changes we made to the surveys as a result of our testing process in Section 5 below and in subsequent papers. However, in order to elucidate for the reader what, exactly, each survey contains, we use this section to describe the content of the final versions of the four surveys. See Appendices A-D for copies of these surveys.

4.1. Form A

There are six items on Form A, all of which relate to the *Hubble plots* construct. Items 1-4 ask students to select one or more graphs from a bank of eight Hubble plots that correspond to a given situation (constant expansion, constant contraction, accelerating expansion, and decelerating expansion). Items 5 and 6 both refer to a provided Hubble plot (of a universe expanding at a constant rate, although it is not identified as such). Item 5 asks students to draw a new Hubble plot for a universe that expands twice as fast as the given Hubble plot. Item 6 asks students to draw a Hubble plot for a universe that took much longer than the universe depicted by the given Hubble plot to reach its current size. All six questions require students to explain their reasoning behind their choices, so simply choosing or drawing a correct graph is insufficient to be placed high on our construct map. Since the *Hubble plots* construct focuses on students' abilities to reason about the age and expansion rate of the universe using Hubble plots, the six items on this survey form are judged as adequately covering this construct's domain as defined Table 1.

4.2. Form B

Form B has seven items on the *models* construct. Three of these items (Items 1, 2, and 7) ask students to explain what the expansion of the universe means, what the Big Bang means, and what is expanding in the universe, respectively. The other three items (Items 3-6) epitomize generative questions. Item 3 gives students a drawing of our observable universe and asks them to describe what inhabitants of a galaxy at the edge of our observable universe would see if they looked at regions beyond Earth's observable universe. This item probes whether or not students think there are galaxies beyond our observable universe. This helps identify whether they believe the universe itself has an edge defined by the edge of our observable universe. Item 4 probes whether or not students think the Big Bang was an (explosive) event located in empty space. Item 5 addresses whether or not the universe has a center. Item 6 asks if there is an edge to the distribution of galaxies in the universe (which, as we found in previous surveys, many students do think, especially if they conceive of the Big Bang as "throwing" galaxies out into empty space). We made a deliberate decision on Items 3-6 to not simply ask students if, for example, the universe has a center, because many students can simply regurgitate an answer they have heard in class even if it is not integrated into their fundamental conceptualizations of the expansion of the universe (what Vosniadou 1994 calls "inert knowledge"). We believed these questions probe areas of potential difficulty revealed by previous research and cover the construct's domain as listed in Table 2.

4.3. Form C

Form C focuses on the *evolving universe* construct. It includes six items. Items 1-3 ask students if the temperature, total amount of matter, and density of matter have changed over time and to explain their reasons for their answers. Items 4-6 of Form C all address the relationship between expansion, distances, light travel times, and lookback times. Once again, the items on this survey were chosen based on the information we wanted to learn about students' understanding on this construct. These items both span the construct and help place students on the construct map in Table 3.

4.4. Form D

Form D measures students' knowledge on the *dark matter* construct. Items 1 and 2 both refer to a bank of six different galaxy rotation curves. Item 1 asks which galaxy rotation curve corresponds to a solar system and Item 2 asks which galaxy rotation curve corresponds to a spiral galaxy. The next two items (Items 3 and 4) ask students to rank the relative orbital speeds of planets in a solar system and stars in a galaxy, respectively, based on these rotation curves. Two more items (Items 5 and 6) ask students where the majority of mass in a solar system and a spiral galaxy, respectively, are located. The final item on Form D asks students to compare planets orbiting a star and stars orbiting a galaxy and state whether or not the two situations are similar. These items help place students on the *dark matter* construct map and cover the domain of the construct (Table 4).

5. VALIDATION PROCESS

Before we can draw any conclusions from students' responses to our surveys, we must assess the validity of our surveys. As mentioned above, a survey is valid only if it measures what its designers intend it to measure.

Validity is not a property of a test *per se*. Rather, it is associated with the interpretation one gives to test scores (AERA, APA, and NCME 1999; Kane 1992; Wilson 2005). Kane (1992) summarizes this view of validity by explaining how validity involves an interpretive argument:

> "A test-score interpretation always involves an *interpretive argument*, with the test score as a premise and the statements and decisions involved in the interpretation as conclusions. The inferences in the interpretive argument depend on various assumptions, which may be more-or-less credible. [...] Because it is not possible to prove all of the assumptions in the interpretive argument, it is not possible to verify this interpretive argument in any absolute sense. The best that can be done is to show that the interpretive argument is highly plausible, given all available evidence" (p. 527, italics in original).

Kane further recommends that test designers explicitly state the assumptions for which they must find evidence to support their interpretations of the test's scores (Kane 1992).

For the purposes of this study, what interpretations and decisions do we want to make? We want to interpret survey scores such that higher scores imply a greater mastery of the relevant construct. We want our comparisons of pre-instructional and post-instructional scores to tell us what effect, if any, instruction had on students' cosmology knowledge. We want to compare the gains of students who used the new suite of cosmology *Lecture-Tutorials* to students who did not to see if the *Lecture-Tutorials* had any effect. These interpretations rest on the following assumptions:

(1) Each survey adequately covers the construct it is intended to measure.
(2) The students who take the surveys are representative of the target population of Astro 101 students – that is, we can generalize our results.
(3) Astro 101 students correctly read and interpret our survey items.
(4) Students' responses reveal their ideas about cosmology.
(5) Students' responses can be reliably transformed into numerical scores.
(6) These scores can be used to find measurable differences between different populations of students.
(7) Differences in the learning gains of students who have and have not used the cosmology *Lecture-Tutorials* are due to the *Lecture-Tutorials* and not some other variable.

This is a large list, and we cannot support every entry with the same quantity of evidence. To reiterate Kane (1992), we cannot provide any absolute proof of the validity of our surveys. However, we have collected multiple lines of evidence in support of our interpretive argument and, therefore, the validity of our four surveys. In the rest of this section, we focus on qualitative pieces of validity evidence. These cover much of the first four entries in our above list of assumptions. We address the other assumptions with the quantitative evidence presented in future papers in this series.

5.1. Each Survey Adequately Covers the Construct it is Intended to Measure

Much of our design process explicitly addresses this issue. We devoted much of this paper to explaining how we defined our constructs, developed our construct maps, and designed our items precisely because we want to establish the connection between the content of our surveys and the constructs we intend them to measure.

Additionally, multiple experts in the fields of physics, astrophysics and education reviewed our surveys' items. These experts included physics and astronomy education researchers, astronomers, and graduate students at two institutions. Specifically, our survey items were reviewed by members of the Physics Education Research Group at the University of Colorado at Boulder and the Center for Astronomy Education (CAE) at the University of Arizona. These reviewers evaluated a preliminary bank of items we considered for inclusion in the surveys. They evaluated the items along three dimensions. First, they looked at whether or not the items would make sense to an Astro 101 student. Second, they considered whether they would expect an Astro 101 student to be able to answer these items post-instruction. Finally, they examined the items for possible errors in the relevant astrophysics.

They did not comment much on the content of the items, but they did help clarify the wording and presentation of several items. We were able to reduce the total number of items to be used based on their suggestions.
We selected a subset of items that survived this review and used these items for our first drafts of Forms A-D. We subsequently gave these drafts to three graduate students in the School of Education. Their input was valuable since

they are the only people who reviewed the items who do not identify themselves primarily as practicing physicists or astronomers. These three students examined the questions and pointed out word choices that were potentially confusing for non-experts. We revised the surveys based on the recommendations of these three graduate students and then presented the surveys to an expert panel of three astrophysicists drawn from the University of Colorado at Boulder's Department of Astrophysical and Planetary Sciences. All participating astrophysicists conduct research in cosmology, teach cosmology in Astro 101, or both. Again, they looked at whether or not the items would make sense to an Astro 101 student, whether they would expect an Astro 101 student to be able to answer these items post-instruction, and whether the items contained any errors in the relevant astrophysics. Most of their comments resulted in only minor modifications to the items. A follow-up review of the items by astrophysicists at the University of Arizona yielded the same results. We revised the survey items based on the feedback from this process and administered these surveys to students for the first time in the fall of 2009.

5.2. The Students Who Take the Surveys Are Representative of the Target Population of Astro 101 Students

That is, we can generalize our results. Our data for this study came from twenty-one classes taught at thirteen different institutions. These span a range of class sizes and institutional types. Our sample includes classes with more than 600 students, classes with fewer than 10 students, and classes with enrollments in between these numbers. It includes community college classes, classes taught at liberal arts colleges, and classes taught at large research-focused institutions. We have both public and private colleges in our sample as well. See Table 5 for a summary of the demographic information for each class in our study.

For this study, we surveyed a total of 2318 students pre-instruction and 2041 students post-instruction. 1709 of the pre-instruction responses and 1527 of the post-instruction responses came from students in classes that used the new cosmology *Lecture-Tutorials*. 609 of the pre-instruction responses and 514 of the post-instruction responses came from students in classes that did not use the new cosmology *Lecture-Tutorials*. The fact that students in all of these varied classes repeatedly exhibited the same set of difficulties with cosmology (see the data in Paper 4 and in Wallace 2011) suggest that the results of this study apply to the broader population of Astro 101 students in the United States.

Table 5. Demographic information for each surveyed class

Class	Semester	State	Public or Private	Used LTs?	Pre-instruction responses	Post-instruction responses
Class A	F09	AZ	Public	Yes	282	231
Class B	F09	CO	Public	Yes	119	100
Class C	F09	CO	Public	No	100	75
Class D	S10	AZ	Public	Yes	687	626
Class E	S10	CO	Public	Yes	237	220
Class F	S10	CO	Public	No	136	86
Class G	S10	NY	Private	No	155	149
Class H	F10	CO	Public	Yes	110	92
Class I	F10	CA	Public	Yes	52	67
Class J	F10	CA	Public	Yes	20	20
Class K	F10	NY	Public	Yes	21	16
Class L	F10	NV	Public	Yes	8	4
Class M	F10	KS	Public	Yes	34	30
Class N	F10	KS	Public	Yes	9	7
Class O	F10	MD	Public	Yes	10	8
Class P	F10	CA	Public	Yes	120	106
Class Q	F10	CO	Public	No	80	65
Class R	F10	CA	Public	No	57	66
Class S	F10	NV	Public	No	6	4
Class T	F10	WI	Public	No	57	53
Class U	F10	MI	Private	No	18	16

5.3. Astro 101 Students Correctly Read and Interpret Our Surveys' Items

To help address this issue, we interviewed nineteen Astro 101 students at the University of Colorado at Boulder over the course of the fall 2009, spring 2010, and fall 2010 semesters. All nineteen students volunteered for an interview by checking a box on the consent forms they signed when they agreed to participate in this study. They received no compensation for their time.

We interviewed these nineteen students in order to get a sense of whether or not Astro 101 students interpreted our survey items as we intended. The interviews were semi-structured think-alouds (Otero and Harlow 2009): We gave each student one survey item at a time and asked him/her to describe everything he/she thought of while constructing his/her answer (Willis 2005). Before the students tried thinking aloud on the survey items, we gave them an unrelated question on which to practice thinking aloud (Otero and Harlow 2009; Willis 2005). We also followed much of Patton's (1980) advice: We avoided dichotomous and "why" questions, we had one idea per question, and we made sure that the student did the lion's share of the talking during the interview. We prepared several follow-up questions to each item, in case a student did not adequately explain a key point of interest (Willis 2005).

We videotaped all nineteen interviews. We also took notes during the interviews and recorded a summary and impressions of each interview immediately after the interviewee left (Erickson 1986; Patton 1980).

Table 6 lists the interviewed students by pseudonym. It shows in which semester they were interviewed, which surveys we used during their interview, and in what order those items were presented. For example, we first interviewed Melissa on the items on Form B, followed by Form A, and ending with Form D. We varied the order of the items on the surveys because students tend to give poorer quality responses to items asked early in an interview. This is because many students are not used to participating in a think-aloud interview, and begin to offer more detailed responses as the interview progresses.

These interviews helped us detect several issues with the wordings of certain items in the fall 2009 and spring 2010 versions of the surveys. To give just one example, Item 4 on the spring 2010 version of Form B contained the phrase "surrounding the event called the Big Bang." Multiple interviewed students expressed confusion about the term "surrounding the event." Brenda's response encapsulates this confusion:

> "Surrounding the event? Like is this, I mean (sighs), this question is confusing. Surrounding the event, like of the Big Ba, the Big Bang, the time when they thought it actually happened, like the moment of

Table 6. The forms each student responded to during his/her interview. The numbers denote the order in which we presented the surveys to the student

Student	Semester	Form A	Form B	Form C	Form D
Melissa	F09	2	1	–	3
Tyson	F09	3	2	1	–
Kelsey	F09	–	3	2	1
Abigail	F09	1	2	3	–
Nina	S10	1	3	–	2
John	S10	3	–	1	2
Paul	S10	–	1	2	3
Gayle	S10	2	–	3	1
Calvin	S10	3	2	1	–
Brenda	S10	2	1	3	–
Molly	F10	2	–	3	1
Patrick	F10	3	2	–	1
Eduardo	F10	2	–	1	3
Stan	F10	3	–	1	2
Vanessa	F10	1	3	–	2
Cecelia	F10	–	1	3	2
Timothy	F10	–	1	2	3
Tucker	F10	1	2	3	–
Brett	F10	1	3	2	–

the Big Bang, or like right before or right after, like surrounding the event? Or s- like, I don't know what this question's asking. I mean I guess I could assume that it means surrounding like the moment of the Big Bang or, like the moment prior, like, immediately prior to would be, like, a region of space that includes nothing because everything that was in the universe was, like, in this one little space and then expanded. I don't know. That question's super confusing."

This item was completely rewritten because of responses such as Brenda's. We made significant revisions to all four forms after each semester as a result of the feedback we received from these interviews.

In the fall of 2010, none of our nine interviewed students found any problems with any of the items on any of the surveys. This suggests that the revisions we made after previous semesters produced surveys that are comprehensible in their entireties to Astro 101 students.

5.4. Students' Responses Reveal Their Ideas About Cosmology

Our survey items elicit a wide range of student ideas about the Big Bang, the expansion and evolution of the universe, and dark matter. For example, consider two pre-instruction responses to the first item on Form B: "Explain, in as much detail as possible, what astronomers mean when they say 'the universe is expanding.' Provide a drawing if possible to help illustrate your thinking." One student wrote the following:

> "When astronomers say the universe is expanding they are referencing that all stars and planets are moving out away from a central area in the universe."

This student included with his/her response the drawing shown in Figure 1, along with the note that each arrow "indicates movement of stars." In contrast, a different student wrote:

> "[T]here is a raisin bread effect as the dough rises the yeast expands (think of this as space) as this is happening, all of the raisins (our galaxies) get farther + farther appart [*sic*] from each other."

These two students' answers to the same item illustrate how a single item can elicit responses that reveal very different conceptions about the expansion of the universe.

We also saw a wide range of student responses to other survey items. To take just two examples, consider Item 2 from Form B ("Explain, in as much detail as possible, what astronomers mean by the 'Big Bang Theory.' Provide a drawing if possible to help illustrate your thinking.") and Item 1 from Form C ("Over time, would you say the temperature of the universe has increased, decreased, or stayed the same? Explain your reasoning and provide a drawing if possible to help illustrate your thinking."). In response to the first item, one student described the Big Bang as follows:

> "The Big Bang Theory, in a nutshell, says that the universe started out as very hot + dense, and over time that energy was converted into matter as the distance between matter increased and as the universe expanded."

Another student had a very different, non-expert-like idea about the Big Bang:

> "'the big bang theory' is when an asteroid that was headed toward earth struck the earth and every thing that was alive died – then as time went on things started growing and living again."

As for the temperature of the universe (the second item), we received some responses that could be classified as correct, such as the following:

Figure 1. A student's drawing of the expansion of the universe

> "The temp. has gotten cooler. Again assuming no matter is created in the expansion process of the universe than [*sic*] while the universe has expanded it has become less dense and as you become less dense you cool off."

Other responses illustrate how a student can have the right answer, but for the wrong reason. For example, one student wrote:

> "It could be getting colder as time goes on because as the universe expands there is more and more area for the sun and other giant stars to heat, and also the sun is continuously collapsing on itself (I think?) which might mean its getting smaller, just like other giant stars in the universe are which could also be making it cooler as time progresses."

These quotations from actual student answers reveal just a fraction of the diversity of responses we received from the thousands of students that participated in this study. We discuss the prevalence of these and other student ideas in Paper 4. At this point, the quotations shown above illustrate that our surveys can reveal a wide range of ideas about cosmology from Astro 101 students.

At this point, we have addressed evidence for the first four parts of our validity argument:

(1) Each survey adequately covers the construct it is intended to measure.
(2) The students who take the surveys are representative of the target population of Astro 101 students – that is, we can generalize our results.
(3) Astro 101 students correctly read and interpret our survey items.
(4) Students' responses reveal their ideas about cosmology.

As noted above, we will address other parts in subsequent papers in this series.

6. SUMMARY

We have developed four surveys that can be used to investigate Astro 101 students' conceptual understandings of commonly taught cosmology topics. These topics include the expansion and evolution of the universe, the Big Bang, and the evidence for dark matter in spiral galaxies. In this paper, we detailed the constructs each survey measures, the principles that guided the design of our surveys' items, and the content of our surveys.

We also outlined each component of the interpretive argument used to establish the validity of our surveys. This interpretive argument is a reoccurring theme in future papers in this series; each paper adds additional evidence in support of our surveys' validity. In this paper, we presented qualitative evidence that the surveys adequately cover the constructs they are designed to measure, that the students who participated in this study are representative of the broader Astro 101 population, that students correctly read and interpret survey items, and that students' responses to the items reveal their ideas about cosmological topics.

Papers 2 and 3 use classical test theory and item response theory, respectively, to quantitatively analyze students' survey responses. The data in these papers further illustrates the evolution of our surveys, and they address more quantitative aspects of our validity argument, such as whether or not students' responses can be reliably transformed into numerical scores.

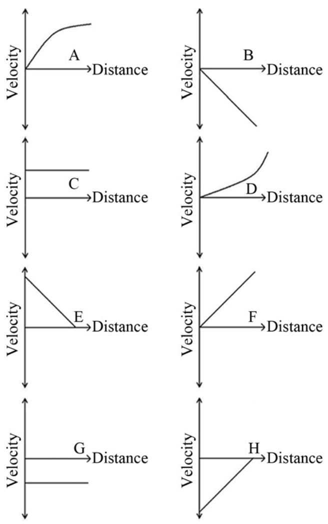
This paper did not answer this study's underlying research questions. This paper, plus Papers 2 and 3, set the stage for Papers 4 and 5, which explore the nature of students' difficulties with cosmology and the efficacy of the new suite of cosmology *Lecture-Tutorials*, respectively.

Acknowledgments

This work would not have been possible without the assistance of the Physics Education Research Group at the University of Colorado at Boulder and members of the Center for Astronomy Education (CAE) at the University of Arizona. We also thank the members of the School of Education and the Department of Astrophysical and Planetary Sciences at the University of Colorado at Boulder, as well as the members of the Department of Astronomy at the University of Arizona who reviewed our survey items. We owe a special thanks Derek Briggs, from whom we learned the process of survey design and validation. Gina Brissenden and Sébastien Cormier provided valuable feedback on earlier versions of this manuscript. This study would not have been possible without the many instructors and students who volunteered to participate. This paper benefited from the anonymous reviewer who suggested edits that improved the final version of this paper. The image of the galaxy in Appendix D is courtesy of R. Gendler. This material is based in part upon work supported by the National Science Foundation under Grant Nos. 0833364 and 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 authors and do not necessarily reflect the views of the National Science Foundation.

Appendix A: Form A—Final Version (Fall 2010)

Below are eight graphs (A–H) showing how fast galaxies are moving (velocity) versus their distances away from us.



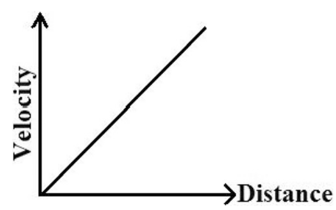
(1) Which graph or graphs (A–H), if any, show a universe that is expanding at a constant rate? Explain your reasoning for your selection(s). If your answer is "none," explain why.

(2) Which graph or graphs (A–H), if any, show a universe that is contracting at a constant rate? Explain your reasoning for your selection(s). If your answer is "none," explain why.

(3) Which graph or graphs (A–H), if any, show a universe that is expanding at a faster and faster rate over time? Explain your reasoning for your selection(s). If your answer is "none," explain why.

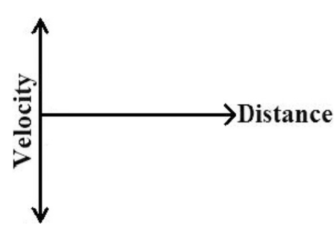
(4) Which graph or graphs (A–H), if any, show a universe that is expanding at a slower and slower rate over time? Explain your reasoning for your selection(s). If your answer is "none," explain why.

Figure 1, below, is a possible graph showing how fast galaxies move away from us in the expanding universe.



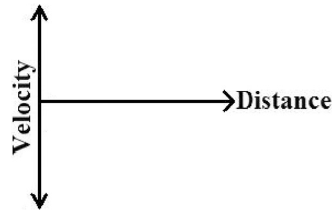
(5) Use the blank graph provided below to draw what you think Figure 1 would look like if the universe had been expanding twice as fast. Explain the reasoning behind the graph you drew.

If you don't have enough information to do this, explain what else you need to know.



(6) Use the blank graph provided below to draw what you think Figure 1 would look like for our universe if it took much longer to reach its current size. Explain the reasoning behind the graph you drew.

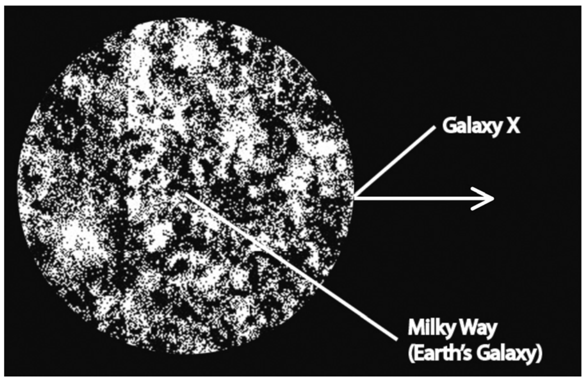
If you don't have enough information to do this, explain what else you need to know.



Appendix B: Form B—Final Version (Fall 2010)

(1) Explain, in as much detail as possible, what astronomers mean when they say "the universe is expanding." Provide a drawing if possible to help illustrate your thinking.

(2) Explain, in as much detail as possible, what astronomers mean by the "Big Bang Theory." Provide a drawing if possible to help illustrate your thinking.



(3) Each white dot in the picture above is a galaxy. The Milky Way Galaxy (the one we live in) is at the center of the picture. All of the galaxies inside the circle can be seen from Earth. Any galaxies that exist outside the circle are so far away that their light has not had time to reach Earth. Describe what inhabitants of Galaxy X probably see when they look in the direction of the arrow.

(4) Circle the sentence that best describes the universe at the time of the Big Bang. Explain your reasoning for your choice and provide a drawing if possible to help illustrate your thinking.

(a) In the beginning, there was space in the universe surrounding the location of the Big Bang but this space was empty of all matter.

(b) In the beginning, there was space in the universe surrounding the location of the Big Bang and matter already existed in this space.

(c) I think of the Big Bang differently than a or b.

(5) Independent of whether we know its true location, is there a center to the universe? Explain your reasoning and provide a drawing if possible to help illustrate your thinking.

(6) If you could travel to any location in the universe, could you go to a place where there are no galaxies in front of you (a.k.a. empty space)? Explain your reasoning and provide a drawing if possible to help illustrate your thinking.

(7) Which of the following statements [(a)–(d)] are true? Circle all that apply.
In general, the expansion of the universe causes _________.

(a) the distances between <u>planets in the solar system</u> to increase.

(b) the distances between <u>stars in the galaxy</u> to increase.

(c) the distances between <u>galaxies in the universe</u> to increase.

(d) None of the above.

Explain your reasoning for your choice(s).

Appendix C: Form C—Final Version (Fall 2010)

(1) Over time, would you say the temperature of the universe has increased, decreased, or stayed the same? Explain your reasoning and provide a drawing if possible to help illustrate your thinking.

(2) How does the total amount of matter (not energy) in the universe *right now* compare to the total amount of matter at the *very beginning* of the universe (the moment just after the Big Bang)? Explain your reasoning and provide a drawing if possible to help illustrate your thinking.

(3) Over time, would you say the overall density of matter in the universe has increased, decreased, or stayed the same? Explain your reasoning and provide a drawing if possible to help illustrate your thinking.

Questions 4–6 refer to this situation: Galaxy X and Galaxy Y are part of an expanding universe. Inhabitants of Galaxy X see a star explode in Galaxy Y. They determine that the light from the explosion took exactly 8 billion years to reach Galaxy X.

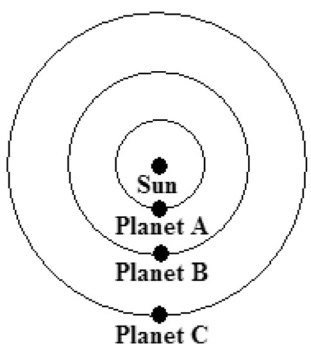
(4) How far apart were Galaxy X and Galaxy Y when the star exploded in Galaxy Y?
a) less than 8 billion light-years apart b) exactly 8 billion light-years apart
c) more than 8 billion light-years apart d) there is not enough information to tell
Explain your reasoning for your choice and provide a drawing if possible to help illustrate your thinking.

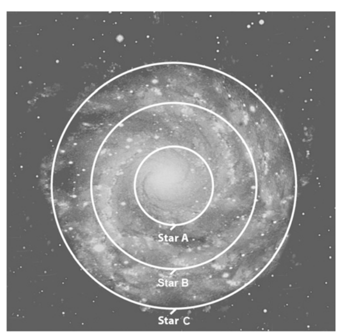
(5) How far apart are Galaxy X and Galaxy Y when Galaxy X sees the star explode?
a) less than 8 billion light-years apart b) exactly 8 billion light-years apart
c) more than 8 billion light-years apart d) there is not enough information to tell
Explain your reasoning for your choice and provide a drawing if possible to help illustrate your thinking.

(6) The universe is 13 billion years old when inhabitants of Galaxy X see the star in Galaxy Y explode. When these inhabitants look at the exploding star, they see what Galaxy Y was like when the universe is what age?
a) less than 5 billion years old b) exactly 5 billion years old
c) more than 5 billion years old d) there is not enough information to tell
Explain your reasoning for your choice and provide a drawing if possible to help illustrate your thinking.

Appendix D: Form D—Final Version (Fall 2010)

On the left is a picture of a solar system with three planets (A–C). On the right is a picture of a spiral galaxy which shows the locations of three stars (A–C).



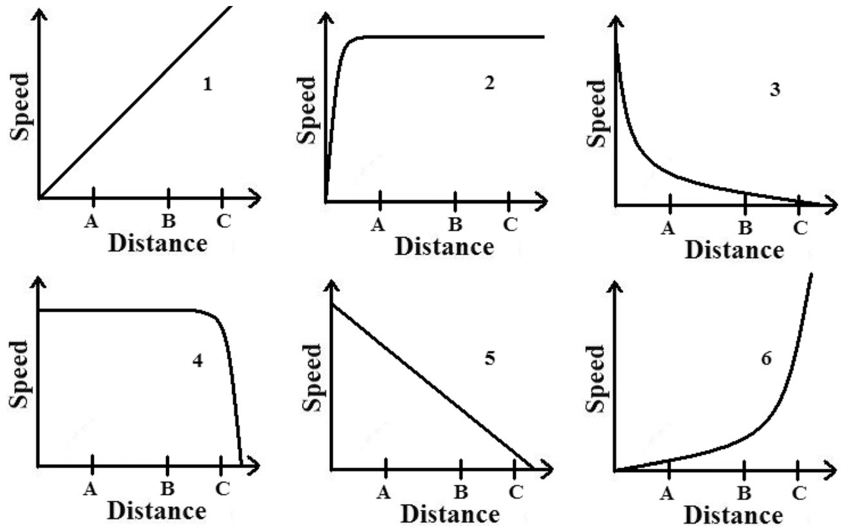


Astronomers make plots called rotation curves when they observe how fast planets orbit the Sun versus their distance from the Sun. They also plot rotation curves when they observe how fast stars orbit the center of a galaxy versus their distance from the center of the galaxy. Below are some possible rotation curves (1–6).

(1) Which graph (1–6) best represents how **planets** orbit the Sun?

(2) Which graph (1–6) best represents how **stars** orbit the center of the galaxy?

(3) Rank the speeds at which **Planets** A, B, and C orbit the Sun:
Ranking Order: Fastest speed 1________ 2________ 3________ Slowest
Or, all the planets orbit at approximately the same speed. _____ (indicate with a check mark)



Explain your reason for ranking this way:

(4) Rank the speeds at which **Stars** A, B, and C orbit the galaxy.
Ranking Order: Fastest speed 1________ 2________ 3________ Slowest
Or, all the stars orbit at approximately the same speed. _____ (indicate with a check mark)
Explain your reason for ranking this way:

(5) Based on your previous answers, how is matter distributed in **solar systems**? Pick the best answer from the following choices [(a)–(c)].
(a) Most of the matter in the solar system is located in the Sun.
(b) Most of the matter in the solar system is evenly distributed throughout the Sun and planets.
(c) My thinking is different than a and b.
Explain your reasoning.

(6) Based on your previous answers, how is matter distributed in **spiral galaxies**? Pick the best answer from the following choices [(a)–(c)].
(a) Most of the matter in the galaxy is located in the center.
(b) Most of the matter in the galaxy is located in the spiral arms.
(c) My thinking is different than a and b.
Explain your reasoning.

(7) Based on your answers for Questions 1-7, do **stars** orbiting the center of a galaxy act like **planets** orbiting the Sun? If yes, explain why. If no, explain why not.

References

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). 1999, *Standards for Educational and Psychological Testing*, Washington, DC: American Education Research Association.

Bailey, J. M. 2007, "Development of a Concept Inventory to Assess Students' Understanding and Reasoning Difficulties About the Properties and Formation of Stars," *Astronomy Education Review*, 6, 133.

Bailey, J. M., Coble, K., Cochran, G., Sanchez, R., Larrieu, D., Hayes, V. L., Nickerson, V., McLin, K. M., and Cominsky, L. R. 2011. "Dark Energy is 'Dying,' and Other Student Ideas About Cosmology," Invited Talk at the 2011 Winter Meeting of the American Association of Physics Teachers (Jacksonville, FL).

Bailey, J. M., Slater, S. J., and Slater, T. F. 2011, "Test of Astronomy Standards," in *Conducting Astronomy Education Research: A Primer*, New York, NY: W. H. Freeman and Company, 107.

Bardar, E. M., Prather, E. E, Brecher, K., and Slater, T. F. 2007, "Development and Validation of the Light and Spectroscopy Concept Inventory," *Astronomy Education Review*, 5, 103.

Bennett, J., Donahue, M., Schneider, N., and Voit, M. 2008, *The Cosmic Perspective*, 5th ed., San Francisco, CA: Pearson Education, Inc.

Briggs, D. C., Alonzo, A. C., Schwab, C., and Wilson, M. 2006, "Diagnostic Assessment With Ordered Multiple-Choice Items," *Educational Assessment*, 11, 33.

Chaisson, E., and McMillan, S. 2011, *Astronomy Today*, 7th ed., San Francisco, CA: Pearson Education, Inc.

Coble, K., McLin, K. M., Metevier, A. J., Bailey, J. M., and Cominsky, L. R. 2011. "The Big Ideas in Cosmology: A Curriculum for College Students," Invited Talk at the 2011 Winter Meeting of the American Association of Physics Teachers (Jacksonville, FL).

Comins, N. F. 2001, *Heavenly Errors: Misconceptions About the Real Nature of the Universe*, New York, NY: Columbia University Press.

Deming, G., and Hufnagel, B. 2001, "Who's Taking ASTR 101?," *Physics Teacher*, 39, 368.

diSessa, A. A. 1993, "Toward an Epistemology of Physics," *Cognition and Instruction*, 10, 105.

Davis, T M., and Lineweaver, C. H. 2004, "Expanding Confusion: Common Misconceptions of Cosmological Horizons and the Superluminal Expansion of the Universe," *Publications of the Astronomical Society of Australia*, 21, 97.

Erickson, F. 1986, "Qualitative Methods in Research on Teaching," in *Handbook of Research on Teaching*, 3rd ed., ed. M. C. Wittrock, New York, NY: Macmillan Publishing Company, 119.

Gorin, J. S. 2006, "Test Design with Cognition in Mind," *Educational Measurement: Issues and Practice*, 25, 21.

Hansson, L., and Redfors, A. 2006, "Swedish Upper Secondary Students' Views of the Origin and Development of the Universe," *Research in Science Education*, 36, 355.

Henriques, L., Colburn, A., and Ritz, W. C. 2006, "Developing Assessment Items: A How-to Guide," in *Assessment in Science: Practical Experiences and Education Research*, ed. M. McMahon, P. Simmons, R. Sommers, D. DeBaets, and F., Crawley, Arlington, VA: NSTA Press, 15.

Heron, P. R. L. 2004, "Empirical Investigations of Learning and Teaching, Part I: Examining and Interpreting Student Thinking," in *Proceedings of the International School of Physics "Enrico Fermi," Course CLVI*, ed. E. F. Redish and M. Vicentini, Amsterdam, NL: IOS Press, 341.

Hufnagel, B. 2002, "Development of the Astronomy Diagnostic Test," *Astronomy Education Review*, 1, 47.

Kane, M. T. 1992, "An Argument-Based Approach to Validity," *Psychological Bulletin*, 112, 527.

Keller, J. M. 2006, "Part I: Development of a Concept Inventory Addressing Students' Beliefs and Reasoning Difficulties Regarding the Greenhouse Effect," Ph.D. dissertation, University of Arizona.

Lightman, A. P., and Miller, J. D. 1989, "Contemporary Cosmological Beliefs," *Social Studies of Science*, 19, 127.

Lightman, A. P., Miller, J. D., and Leadbeater, B. J. 1987, "Contemporary Cosmological Beliefs," in *Misconceptions and Educational Strategies in Science and Mathematics*, Vol. III, ed. J. D. Novak, Ithaca, NY: Cornell University Press, 309.

Lindell, R. S. 2001, "Enhancing College Students' Understanding of Lunar Phases," Ph.D. dissertation, University of Nebraska.

Lineweaver, C. H., and Davis, T. M. 2005, "Misconceptions About the Big Bang," *Scientific American*, 292, 36.

McDermott, L. C., Rosenquist, M. L., and van Zee, E. H., 1987, "Student Difficulties in Connecting Graphs and Physics: Examples from Kinematics," *American Journal of Physics*, 55, 503.

Otero, V. K., and Harlow, D. B. 2009, "Getting Started in Qualitative Physics Education Research," in *Reviews in PER Vol. 2: Getting Started in PER*, ed. C. Henderson and K. A. Harper, College Park, MA: American Association of Physics Teachers, 1.

Patton, M. Q. 1980, *Qualitative Evaluation Methods*, Thousand Oaks, CA: SAGE Publications.

Prather, E. E., Slater, T. F., and Offerdahl, E. G. 2002, "Hints of a Fundamental Misconception in Cosmology," *Astronomy Education Review*, 1, 28.

Rudolph, A. L., Prather, E. E., Brissenden, G., Consiglio, D., and Gonzaga, V. 2010, "A National Study Assessing the Teaching and Learning of Introductory Astronomy. Part II: The Connection between Student Demographics and Learning," *Astronomy Education Review*, 9, 010107.

Sadler, P. M., Coyle, H., Miller, J. L., Cook-Smith, N., Dussault, M., and Gould, R. R. 2010, "The Astronomy and Space Science Concept Inventory: Development and Validation of Assessment Instruments Aligned with the K-12 National Science Standards," *Astronomy Education Review* 8, 010111.

Seeds, M. A., and Backman, D. E. 2011, *Foundations of Astronomy*, 11th ed., Boston, MA: Brooks/Cole.

Shepard, L. A. 1993, "Evaluating Test Validity," in *Review of Research in Education*, Vol. 19, ed. L. Darling-Hammond, Washington, DC: American Educational Research Association, 405.

Simonelli, G., and Pilachowski, C. A. 2003, "First-Year College Students' Ideas About Astronomy: A Pilot Study," *Astronomy Education Review*, 2, 166.

Slater, T., Adams, J. P., Brissenden, G., and Duncan, D. 2001, "What Topics Are Taught in Introductory Astronomy Courses?," *Physics Teacher*, 39, 52.

Spergel, D. N., Bean, R., Doré, O., Nolta, M. R., Bennett, C. L., Dunkley, J., Hinshaw, G., Jarosik, N., Komatsu, E., Page, L., Peiris, H. V., Verde, L., Halpern, M., Hill, R. S., Kogut, A., Limon, M., Meyer, S. S., Odegard, N., Tucker, G. S., Weiland, J. L., Wollack, E., and Wright, E. L. 2007, "Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology," *Astrophysical Journal Supplement*, 170, 377.

Trowbridge, D. E., and McDermott, L. C. 1980, "Investigation of Student Understanding of the Concept of Velocity in One Dimension," *American Journal of Physics*, 48, 1020.

Trowbridge, D. E., and McDermott, L. C. 1981, "Investigation of Student Understanding of the Concept of Acceleration in One Dimension," *American Journal of Physics*, 49, 242.

Trumper, R. 2000, "University Students' Conceptions of Basic Astronomy Concepts," *Physics Education*, 35, 9.

Vosniadou, S. 1994, "Capturing and Modeling the Process of Conceptual Change," *Learning and Instruction*, 4, 45.

Wallace, C. S. 2011, "An Investigation into Introductory Astronomy Students' Difficulties with Cosmology and the Development, Validation, and Efficacy of a New Suite of Cosmology Lecture-Tutorials," Ph.D. dissertation, University of Colorado at Boulder.

Wallace, C. S., Prather, E. E., and Duncan, D. 2011a, "A Study of General Education Astronomy Students' Understandings of Cosmology. Part II. Evaluating Four Conceptual Cosmology Surveys: A Classical Test Theory Approach," *Astronomy Education Review*, 10, 010107.

Wallace, C. S., Prather, E. E., and Duncan, D. 2011b, "A Study of General Education Astronomy Students' Understandings of Cosmology. Part III. Evaluating Four Conceptual Cosmology Surveys: An Item Response Theory Approach," *Astronomy Education Review* (submitted).

Wallace, C. S., Prather, E. E., and Duncan, D. 2011c, "A Study of General Education Astronomy Students' Understandings of Cosmology. Part IV. Common Difficulties Students Experience with Cosmology," *Astronomy Education Review* (submitted).

Wallace, C. S., Prather, E. E., and Duncan, D. 2011d, "A Study of General Education Astronomy Students' Understandings of Cosmology. Part V. The Effects of a New Suite of Cosmology Lecture-Tutorials on Students' Conceptual Knowledge," (in preparation).

Willis, G. D. 2005, *Cognitive Interviewing: A Tool for Improving Questionnaire Design*, Thousand Oaks, CA: SAGE Publications.

Wilson, M. 2005, *Constructing Measures: An Item Response Modeling Approach*, Mahwah, NJ: Lawrence Erlbaum Associates.

ÆR
010106-1–010106-20