International Journal of Science Education
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tsed20

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

Colin S. Wallace [a], Edward E. Prather [a] & Douglas K. Duncan [b]
[a] Center for Astronomy Education (CAE), Steward Observatory, University of Arizona, Tucson, AZ, USA
[b] Department of Astrophysical and Planetary Sciences, University of Colorado at Boulder, Boulder, CO, USA

Available online: 23 Apr 2012

International Journal of Science Education
2012, 1–18, iFirst Article

Routledge
Taylor & Francis Group

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

Colin S. Wallace[a*], Edward E. Prather[a] and Douglas K. Duncan[b]
[a]*Center for Astronomy Education (CAE), Steward Observatory, University of Arizona, Tucson, AZ, USA;* [b]*Department of Astrophysical and Planetary Sciences, University of Colorado at Boulder, Boulder, CO, USA*

This is the final paper in a five-paper series describing our national study of the teaching and learning of cosmology in general education astronomy college-level courses. A significant portion of this work was dedicated to the development of five new *Lecture-Tutorials* that focus on addressing the conceptual and reasoning difficulties that our research shows students have with frequently taught cosmology topics, such as the expansion of the universe, the Big Bang, and dark matter. We conducted a systematic investigation of the implementation of these new *Lecture-Tutorials* and resulting learning gains in order to test the efficacy of these new *Lecture-Tutorials*. Our investigation included classroom observations, results from pre–post testing using four conceptual cosmology surveys, and comparisons between classes in terms of the class time spent on cosmology topics and other instructional strategies used to teach cosmology. We used this combination of qualitative and quantitative research results to evaluate the conceptual understandings of students who used the new cosmology *Lecture-Tutorials* compared to those students who did not. The analysis of our data shows that, in many cases, classrooms that used the cosmology *Lecture-Tutorials* saw a greater increase in their students' conceptual cosmology knowledge compared to classrooms that did not use the cosmology *Lecture-Tutorials*. However,

*Corresponding author: Center for Astronomy Education (CAE), Steward Observatory, University of Arizona, 933 N Cherry Avenue, Tucson, AZ 85721, USA. Email: cswallace@email.arizona.edu

ISSN 0950-0693 (print)/ISSN 1464-5289 (online)/12/000001–18

http://dx.doi.org/10.1080/09500693.2012.677960

our results also indicate how instructors implement the *Lecture-Tutorials* into their classrooms strongly influences their students' learning gains.

Keywords: *Astronomy; Cosmology; Lecture-Tutorials; Big Bang; Expansion of the universe; Dark matter*

Introduction

General education introductory astronomy courses (hereafter Astro 101) serve a critical role in the science education of many college students in the USA. Each year, up to a quarter of a million college students take an Astro 101 course (Fraknoi, 2002). These students are broadly representative of the nation's overall undergraduate population (Deming & Hufnagel, 2001; Rudolph, Prather, Brissenden, Consiglio, & Gonzaga, 2010). For many of these students, Astro 101 is the last science course they will ever take. Astro 101 thus represents the final opportunity for these students—who will become the US' future politicians, journalists, business leaders, artists, lawyers, as well as teachers, policy makers, voters, and parents—to develop scientific literacy.

One of the most commonly taught topics in Astro 101 is cosmology (Slater, Adams, Brissenden, & Duncan, 2001). By studying cosmology, Astro 101 students can learn how science addresses some of the most fundamental questions asked throughout the history of mankind, such as 'How did the universe begin?' and 'What is the fate of the universe?' Cosmology also provides an engaging topic area in which students can sharpen and extend the reasoning and analysis skills needed to be contributing and successful members of a global society. Yet few studies, to date, have investigated where and why Astro 101 students struggle with cosmology. This project is one of the first large-scale, systematic studies of Astro 101 students' conceptual and reasoning difficulties with cosmology.

This paper is the fifth in a five-paper series detailing the research design, methodologies, analysis, and findings of our national study. The 21 different US Astro 101 courses that participated in this study during the Fall 2009, Spring 2010, and Fall 2010 semesters span a wide range of class sizes (from less than 10 students to over 600) and institution types (including public and private institutions drawn from the US' diverse set of community colleges, liberal arts schools, and PhD-granting research-intensive universities). We asked students from these courses to answer open-response questions from four different survey forms near the start and end of their Astro 101 courses. To provide validation data for students' survey responses, we conducted a series of one-on-one think-aloud interviews with a small sub-set of students. Additionally, we conducted a series of classroom observations to document which active engagement instructional strategies were used and the amount of class time that was dedicated to instruction on the cosmology topics that were the focus of our study. In total, 4,359 student survey responses (pre- and post-instruction) were gathered and analysed. Each of the four different survey forms used for this study probed a different construct (e.g. 'the concept or characteristic that a test is designed to measure'; AERA, APA, & NCME, 1999). The constructs for each survey form (A–D) are as follows:

- Form A: This survey examined students' abilities to interpret Hubble plots.
- Form B: This survey examined students' models of the expansion of the universe and the Big Bang.
- Form C: This survey examined whether or not students understand how the properties of the universe have changed over time.
- Form D: This survey examined whether students could reconstruct the chain of reasoning linking the flat rotation curves of spiral galaxies to the existence of dark matter.

See the first paper in our series (Wallace, Prather, & Duncan, 2011a, aka 'Paper 1') for more demographic details, information on the design of the survey forms, and for copies of the survey forms. Paper 1 also describes the data from the student think-aloud interviews and from feedback from Astro 101 instructors and education researchers that we used to help validate the survey forms.

Paper 2 (Wallace, Prather, & Duncan, 2011b) and Paper 3 (Wallace, Prather, & Duncan, 2012a) describe our classical test theory and item response theory analyses of the data, respectively. As part of these analyses, we developed detailed scoring rubrics that we used to score students' survey responses. Each question was scored on a scale of 0–2 or 0–3, where the maximum score always represented a correct and complete response. By quantifying students' performances on the surveys, we were able to find evidence for the reliability and validity of the survey forms via an inter-rater reliability analysis, Cronbach's α, and Wright maps, and refine the survey forms.

Paper 4 (Wallace, Prather, & Duncan, 2012b) describes the most common pre-instruction conceptual and reasoning difficulties that Astro 101 students experience while studying the cosmology topics we investigated. The difficulties we uncovered informed the design of a new suite of five cosmology *Lecture-Tutorials*. In the current paper, we test the effectiveness of the new *Lecture-Tutorials* by analysing the pre- and post-instruction survey data along with our classroom observations of the instructional strategies used to teach cosmology and the time spent on teaching cosmology.

The cosmology *Lecture-Tutorials* mimic the design of the original *Lecture-Tutorials for Introductory Astronomy* (Prather, Slater, Adams, Brissenden, & Dostal, 2008). Each *Lecture-Tutorial* is a two to six page worksheet comprised of Socratic-style questions. These questions are all related to a single topic, which research has shown to be problematic for students (Prather et al., 2004). Working collaboratively through a *Lecture-Tutorial*'s questions helps students to construct more expert-like understandings of that topic, either by drawing out and building upon students' correct intuitions (Clement, Brown, & Zeitsman, 1989; Elby, 2001) or by making students realize when their intuitions are inappropriate via a conceptual change model commonly characterized by the sequenced phrase 'elicit-confront-resolve' (McDermott, 1991).

Lecture-Tutorials also contain 'student debates': fictionalized arguments between two or more students on a conceptually challenging point. Students working on a *Lecture-Tutorial* must determine if any of the fictionalized students are correct and why (Prather et al., 2004). These student debates act as valuable 'course corrections' for students who have progressed through several questions with their naïve ideas intact. *Lecture-Tutorials* are designed so that they can be easily integrated into the lecture portion of a class (Prather et al., 2004). *Lecture-Tutorials* are explicitly designed to be done collaboratively: each student should work with one or two of her neighbours on a *Lecture-Tutorial*; the majority of the learning that occurs during this time is due to student discourse as they debate and defend their answers and co-construct improved understandings of a topic (Prather et al., 2004). Research shows that students typically achieve larger learning gains using *Lecture-Tutorials* than they do with lecture alone (LoPresto & Murrell, 2009; Prather et al., 2004).

Throughout this paper, we refer to the students and classes that used the new cosmology *Lecture-Tutorials* as 'LT students' and 'LT classes,' respectively. We likewise call the students and classes that did not use the new cosmology *Lecture-Tutorials* as 'non-LT students' and 'non-LT classes,' respectively. We must stress an extremely important point of which the reader must be aware in order to properly interpret our results: 'Non-LT' does not mean 'non-interactive engagement.' We know from first-hand observations and from discussions with the instructors in our study that many 'non-LT' classes consistently used research-validated interactive engagement activities, such as ranking tasks, the original *Lecture-Tutorials for Introductory Astronomy*, and think-pair-share, which is also commonly referred to as Peer Instruction. Each of these instructional activities is known to promote deeper conceptual understandings of topics than traditional, lecture-based instruction (Crouch & Mazur, 2001; Hudgins, Prather, Grayson, & Smits, 2006; Lyman, 1981; Prather, Rudolph, Brissenden, & Schlingman, 2009; Prather et al., 2008). Throughout this paper, the reader must bear in mind that 'non-LT' simply denotes students and classes that did not use the five new cosmology *Lecture-Tutorials* that are the subject of this study.

The remainder of this paper is organized as follows. In the next section, we outline the content of the five cosmology *Lecture-Tutorials*. We then present the results of our analysis of the efficacy of the *Lecture-Tutorials* in improving Astro 101 students' conceptual understandings of cosmology. After discussing our results, we provide the final pieces of evidence in support of the validity of the four conceptual cosmology survey forms (Forms A–D) for this study; this validation argument has been a re-occurring theme throughout the papers in this series. We end with a summary of our results and a discussion of topics for future research.

The Cosmology *Lecture-Tutorials*

We developed five new *Lecture-Tutorials*, each of which focuses on a different conceptually challenging aspect of cosmology (see Paper 4 for more details on students' learning difficulties with cosmology). The five new *Lecture-Tutorials* are:

- 'Dark Matter': We designed this *Lecture-Tutorial* to help students to explain why flat galaxy rotation curves provide evidence for the existence of dark matter.
- 'Hubble's Law': This *Lecture-Tutorial* focuses on students' reasoning difficulties with Hubble's law and helps students learn how to use Hubble plots to infer information about the age and expansion rate of the universe.
- 'Making Sense of the Universe and Expansion': This *Lecture-Tutorial* confronts students' naïve belief that an expanding universe must have a centre and an edge.
- 'Expansion, Lookback Times, and Distances': We designed this *Lecture-Tutorial* to help students to understand how large distances in the universe are related to the concept of lookback time in an expanding universe.
- 'The Big Bang': This *Lecture-Tutorial* helps students to overcome the idea that the Big Bang was an explosion of pre-existing matter into pre-existing empty space.

Note that there is not necessarily a one-to-one correspondence between each *Lecture-Tutorial* and each survey form. While these *Lecture-Tutorials* address the same constructs as the surveys (Forms A–D), we found some constructs to be complex enough that they require multiple *Lecture-Tutorials* in order for students to develop more sophisticated and expert-like understandings.

Results

Do the new cosmology *Lecture-Tutorials* help students to overcome their conceptual and reasoning difficulties with these core topics from cosmology?

In some cases, we found that the LT and the non-LT classes were both quite effective at helping students to overcome their naïve ideas and reasoning difficulties. For example, we found that between 10% and 30% of students, pre-instruction, think 'the expansion of the universe' is simply a metaphor for how our knowledge of the universe grows over time (Paper 4). Another 15% think it is a metaphor for the fact that more objects form in the universe over time (Paper 4). Post-instruction, fewer than 7% of LT and non-LT students maintained either of these ideas.

This result, however, is the exception, not the rule. We found many more conceptual and reasoning difficulties that were overcome more frequently by students who used the new cosmology *Lecture-Tutorials* than by their peers who did not. For example, post-instruction, LT students exhibited fewer naïve ideas about the Big Bang than the non-LT students. As shown in Table 1, the LT population of students was less likely post-instruction to claim that the Big Bang was an explosion and that matter existed before the Big Bang. The LT students were also more likely, post-instruction, to correctly connect the Big Bang to the beginning of the expansion of the universe. We observed improved understandings in the responses of the LT students over the responses of the non-LT students despite the fact that our sampled population was not homogeneous from semester to semester (due to the range of class sizes, institution types, and pedagogical practices represented in each semester's sample).

As another example, we asked students to choose the correct rotation curve for a spiral galaxy from a bank of possible rotation curves (Figure 1). Pre-instruction,

Table 1. Percentage of all students pre-instruction and percentages of LT and non-LT students post-instruction who used the following claims about the Big Bang

Claim	Fall 2009 Pre	Fall 2009 LT post	Fall 2009 Non-LT post	Spring 2010 Pre	Spring 2010 LT post	Spring 2010 Non-LT post	Fall 2010 Pre	Fall 2010 LT post	Fall 2010 Non-LT post
The Big Bang was the beginning of the universe	34	32	21	46	47	46	46	46	73
The Big Bang was the beginning of expansion	12	50	29	10	68	10	20	51	27
The Big Bang was the beginning of something smaller than the universe	16	3	0	22	3	22	10	1	4
The Big Bang was an explosion	53	23	50	52	3	52	56	23	35
Matter existed before the Big Bang	32	15	43	28	6	28	38	26	25

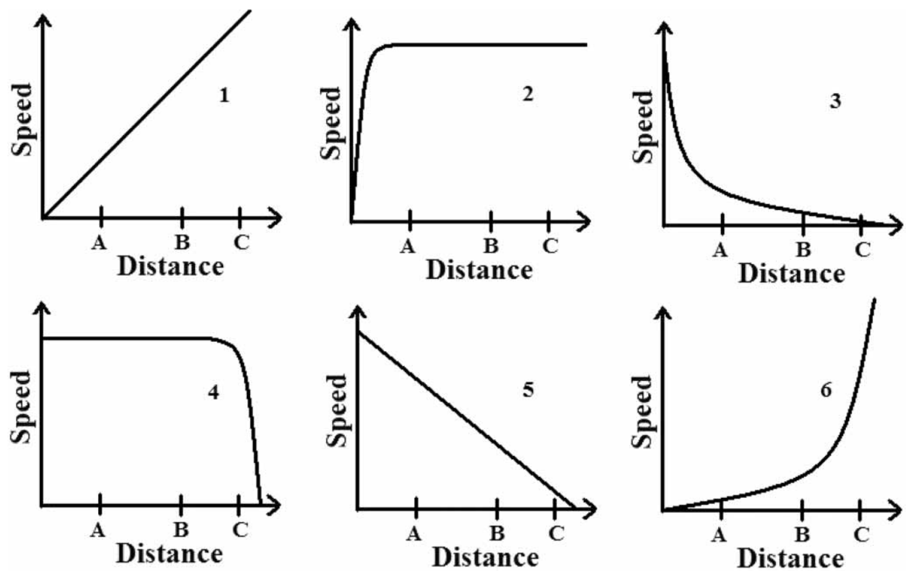

Figure 1. The bank of eight rotation curves from which students had to select the correct rotation curve for a spiral galaxy. Graph 2 is the correct answer

few students chose Graph 2, the correct graph (Figure 2). This result is not surprising—after all, several decades ago even professional astronomers would not have guessed that spiral galaxies have flat rotation curves like Graph 2 in Figure 1. These flat rotation curves are an important piece of evidence for the existence of dark matter. Our post-instruction results on this question are striking. Figure 2

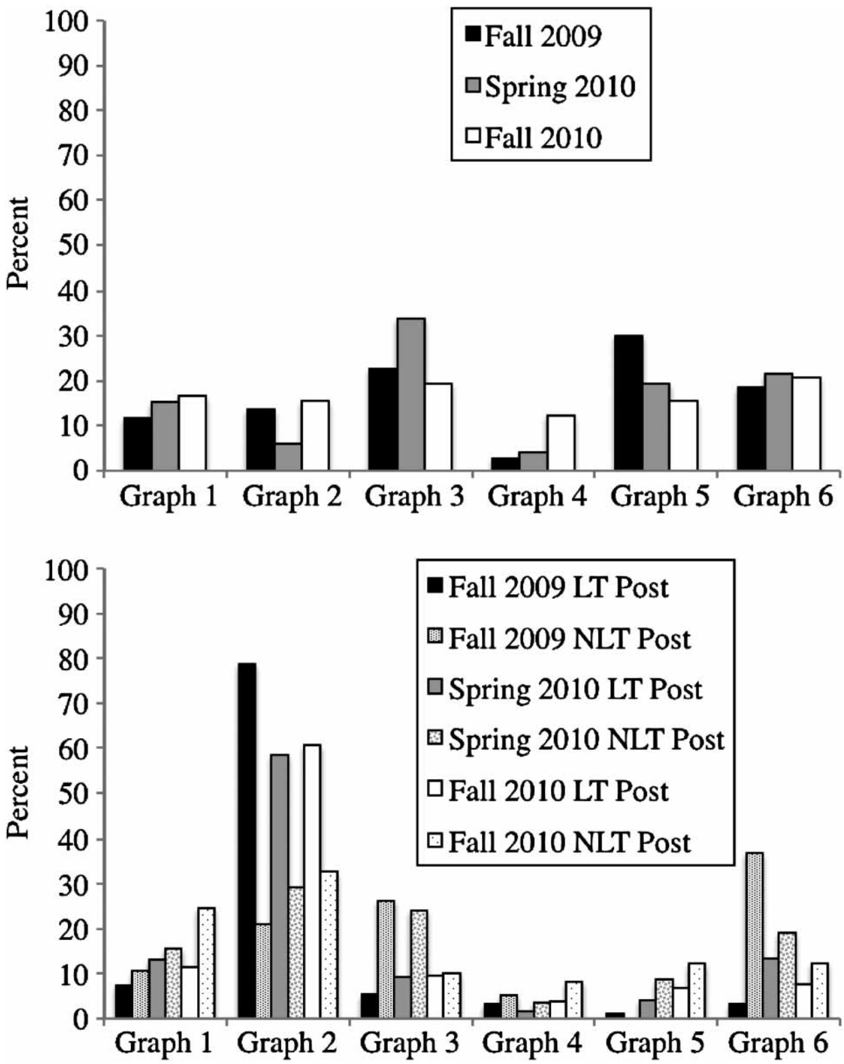

Figure 2. Students' pre-instruction (top) and post-instruction (bottom) choices for the rotation curve of a spiral galaxy. Black bars represent the Fall 2009 responses, grey bars the Spring 2010 responses, and white bars the Fall 2010 responses. In the bottom graph, solid colours correspond to LT students' responses and non-solid colours correspond to non-LT students' responses

shows that LT students were much more likely to recognize the correct rotation curve than their non-LT peers. We observed multiple LT and non-LT classes that provided students with detailed lecture-based instruction on the connection between flat galaxy rotation curves and the arguments used to infer the presence of dark matter. In all these classes, the instruction explicitly emphasized the importance and significance of the flat rotation curve shown in Figure 1 as Graph 2. Yet, the majority of non-LT students were not only unable to identify the correct rotation curve post-instruction—they were also unable to explain the reasoning connecting the shape of the rotation curve with the existence of dark matter (a reasoning pathway probed by subsequent survey items). Working through the 'Dark Matter' *Lecture-Tutorial*, which explicitly requires students to investigate the physical connection between the presence of dark matter and flat rotation curves, demonstrably increased the LT students' ability to identify the correct rotation curve for the orbital motion of objects in the disk of a spiral galaxy.

We also investigated the efficacy of the *Lecture-Tutorials* by videotaping students working through the *Lecture-Tutorials* during their Astro 101 class. Consider, for example, an exchange between two students (pseudonyms Will and Jose) as they finished the 'Dark Matter' *Lecture-Tutorial*. In the following sample of dialogue from the video recording, Will and Jose are discussing the significance of flat rotation curves for our understanding of how matter is distributed in a spiral galaxy:

> Will: So, according to their, like, previous, what they previously thought, if they saw all these things traveling at the same speed, they would have thought they were at the same distance, but they're not so ((inaudible)).
> Jose: Right. Well, they thought that the farther, the farther you got away from the mass –
> Will: The slower the speed.
> Jose: – the slower you would be orbiting, but it turns out what they found is that they were orbiting at the same speed.
> Will: Okay.
> Jose: So that means there's, it's either more evenly distributed from the centre, or there's, or there's, it's, it's like evenly distributed throughout or there's like the mass that we can't see that's messing with the orbit. So there's more mass in the halo than we can see. So we're assuming that all the mass that we're seeing, we're assuming that that's more massive because it's producing more light, but that's not necessarily taking into account the mass that might not be giving off light.

This excerpt provides an example of the kind of critical discourse that a *Lecture-Tutorial* can promote between general education students on conceptually challenging astrophysical problems.

To give the reader a better understanding for the overall performance of each class in our study, Figure 3 shows a plot of the average normalized gain of each class as a function of the class's average pre-test score. The average normalized gain $<g>$ is defined as the difference between a class's average post- and pre-instruction scores (S_f and S_0, respectively), expressed as a fraction of the total amount by which a class's average score could have improved:

$$<g> = \frac{S_f - S_0}{100\% - S_0},$$

(Hake, 1998). There are several key interpretations important to this study that are represented in Figure 3. First, not every class starts with the same average pre-test percentage for a given survey form. This further emphasizes the diversity of classes and students in our sample population. While some classes average 70% prior to instruction, others start out at less than 40%.

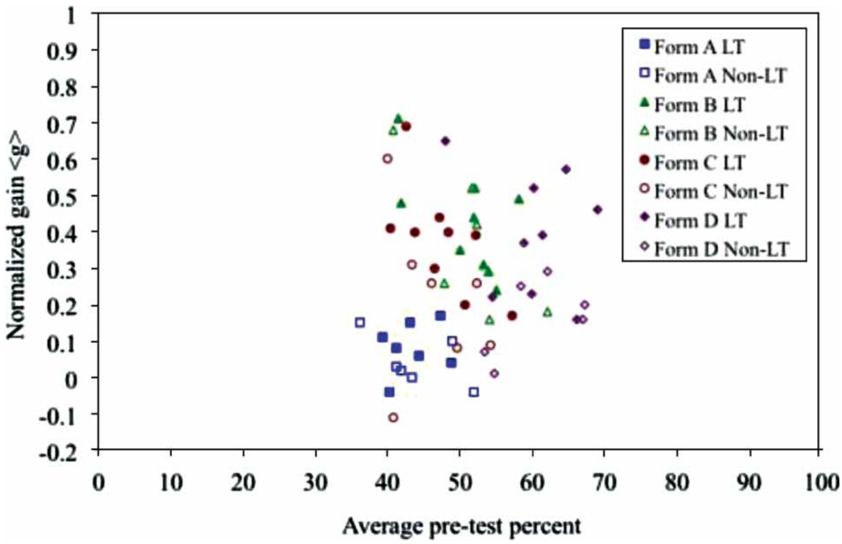

Figure 3. The normalized gains of all LT and non-LT classes as a function of their average pre-test percentages on Forms A–D

Second, a class's average pre-test score does not predict its normalized gain—some classes with low pre-test average scores achieve high gains, while others with high pre-test averages show very little gain. This is important because it shows that one does not need to start out with a high level of conceptual cosmology knowledge in order to do well on our surveys by the end of the course. In other words, Astro 101 students can learn conceptually challenging cosmological topics within a single semester at a level beyond the basic recall of facts.

Third, the data in Figure 3 help us to compare the relative difficulties of the four survey forms. If we only look at pre-instruction scores, then Form A would be categorized as being the most difficult, with an average class pre-instruction percentage of 44%. Form C would be ranked next in difficulty (with a 47% average), followed by Form B (51%), and Form D (60%). If we instead look at normalized gains, the ideas probed by Form A present the greatest challenge to student understanding, with an average class normalized gain of 0.06, followed by Form D (0.30), Form C (0.31), and Form B (0.40). This shows that students enter the Astro 101 classroom with different amounts of pre-instruction knowledge about the constructs probed by the four surveys, and that some constructs (e.g. interpreting Hubble plots) are more difficult for students to master over the course of a semester than others (e.g. models of the Big Bang and the expansion of the universe).

Finally, we see that both LT and non-LT classes are able to improve students' understandings of cosmology topics; however, for a given survey form, the LT classes more often exhibited larger normalized gains than the non-LT classes. If we average the normalized gains for a given survey form across the three semesters, we find that the LT students did better than the non-LT students on all four survey forms (0.08 versus 0.04 on Form A, 0.42 versus 0.37 on Form B, 0.38 versus 0.21 on Form C, and 0.40 versus 0.16 on Form D). Note that all the averages of the non-LT classes, (with the exception of Form B), fall in what Hake (1998) defines as the 'low gain' region ($<g> < 0.30$). In contrast, all the averages of the LT classes, except on Form A, fall within Hake's 'medium gain' region ($0.30 \leq <g> < 0.70$). This supports the hypothesis that students who use the new cosmology *Lecture-Tutorials* develop better conceptual understandings of cosmology than their non-LT peers.

Because the four survey forms changed from semester to semester (see Paper 2), a better comparison is to look at how the LT and non-LT students performed on each survey for each semester. We therefore combined all the LT students together and all the non-LT students together for each survey form for each semester we collected the data. Figure 4 shows the results of this analysis. In almost every case, the LT students achieved larger normalized gains than their non-LT counterparts. Two exceptions to this pattern are the Fall 2010 results for Forms B and C, which we discuss in more detail below. Also, note that neither the LT students nor the non-LT students had large gains on Form A for any semester. The reason for this is also discussed in more detail below.

We next examined the data from both the LT and non-LT groups for each semester in order to determine whether or not the differences between pre- and post-instruction scores are statistically significant. Since students' pre- and post-instruction scores constitute two sets of independent, ordinal data, we used the Mann–Whitney test as our test of statistical significance. We tested our hypothesis that the higher scores are preferentially found in the post-instruction group against the null hypothesis that both high and low scores are equally likely to be found in either the pre- or post-instruction group. Table 2 shows, for each survey form for each semester, whether or not the differences in the pre- and post-instruction scores for the LT and non-LT students are statistically significant ($p < 0.05$). Table 2 shows that these differences are always statistically significant for the LT groups, with the single exception being Form A in the Fall 2009 semester. Table 2 also shows that the differences between the pre- and post-instruction results for the non-LT students are statistically significant in some cases. However, we found that the non-LT results for the Fall 2009 semester were not significant for any of the four survey forms. The results are also not statistically significant for the non-LT students for Form A in the Fall 2010 semester.

Validity

Before we discuss our conclusions, we must finish the validity argument that we have made throughout this five-paper series. We assess the validity of our survey forms to ensure that they are measuring what we think they are measuring. A modern view of validity recognizes it not as a property of a test *per se*, but rather as a function of the interpretation one gives to test scores (AERA, APA, & NCME, 1999; Kane, 1992). Kane (1992) summarizes this view of validity:

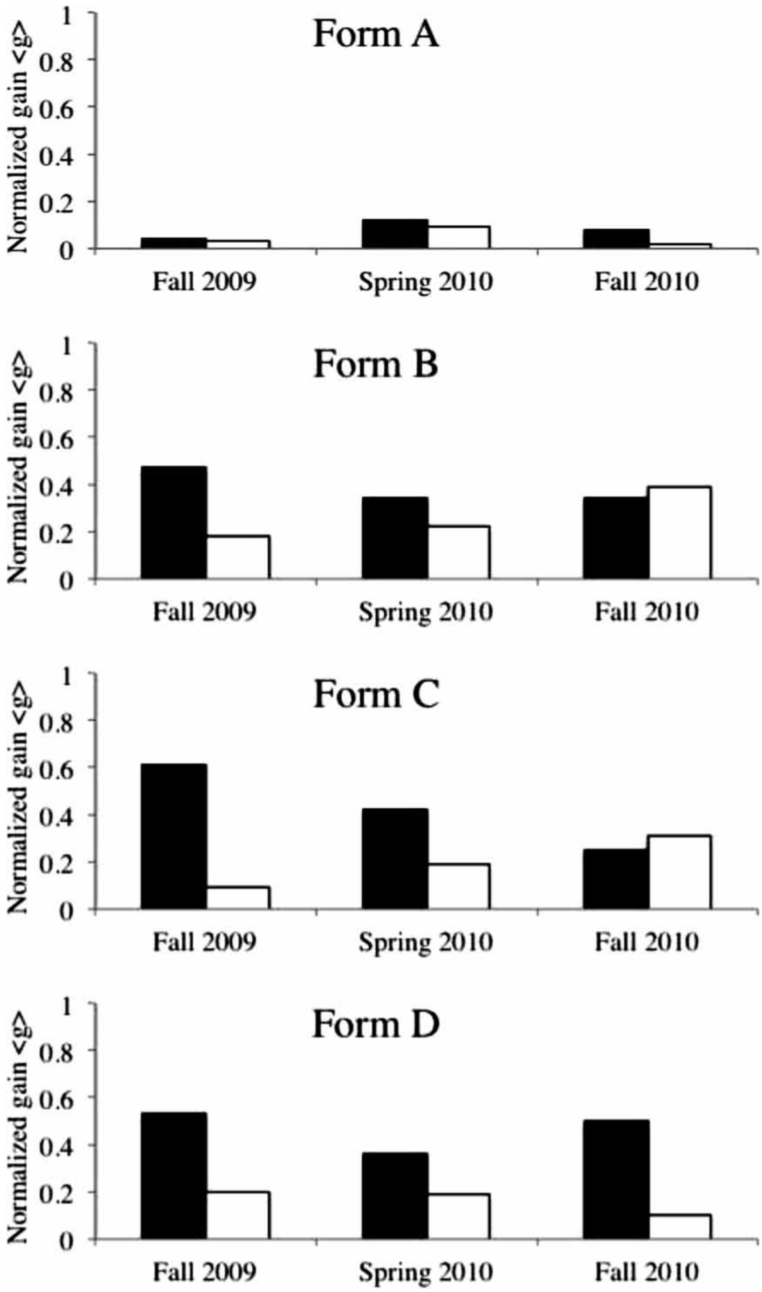

Figure 4. Normalized gains of LT (black bars) and non-LT (white bars) students

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

Table 2. The Mann–Whitney *p*-values for the LT and non-LT groups on Forms A–D for the Fall 2009, Spring 2010, and Fall 2010

Semester	Group	Form A	Form B	Form C	Form D
Fall 2009	LT	0.1003	**<0.0001**	**<0.0001**	**<0.0001**
	Non-LT	0.2514	0.0838	0.1922	0.1230
Spring 2010	LT	**<0.0001**	**<0.0001**	**<0.0001**	**<0.0001**
	Non-LT	**0.0017**	**<0.0001**	**<0.0001**	**0.0275**
Fall 2010	LT	**0.0262**	**<0.0001**	**<0.0001**	**<0.0001**
	Non-LT	0.1762	**<0.0001**	**<0.0001**	**0.0150**

Note: Statistically significant *p*-values ($p < 0.05$) are given in bold.

> 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)

Our interpretive argument is built upon 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.

Previous papers in this series provided evidence for the first five of these assumptions.

We are now in a position to address the last two components of our validity argument: first, can we use students' scores on the survey forms to find measurable differences between different populations of students? Second, to what extent might other variables explain any differences between the results of the LT and non-LT populations?

The results presented in this paper provide the data and analysis we need to answer the first question in the affirmative. With only a few exceptions, we found significant and measurable differences between pre- and post-instruction, and between students that did and did not use the new cosmology *Lecture-Tutorials*.

To answer the second question, we observed four classes, two in the Spring 2010 and two in the Fall 2010. Each semester, we observed one LT class and one non-LT class. Both were Astro 101 courses and both were taught at the same institution, but with different instructors. During each day of class, we took detailed notes on the topics covered, the active engagement instructional methods used (e.g. lecture,

Lecture-Tutorials, think-pair-share, interactive demonstrations, etc.), and the time spent on each topic and method. Following standard procedure for qualitative research (Erickson, 1986), we wrote down our reflections of each lecture immediately after each class or as soon after as possible.

In the Spring 2010, we attended 95% of the LT class's total class time and 92% of the non-LT class's total class time. The LT class devoted 19% of its total class time to cosmology and achieved an average normalized gain (across all four survey forms) of 0.22, while the non-LT class devoted 20% of its total class time on cosmology and achieved an average normalized gain of 0.10. Thus, the LT class did not spend any more time on cosmology and yet achieved a higher average normalized gain.

The Spring 2010 LT class also used more interactive engagement activities than the non-LT class throughout the semester (37% of the LT's class time was spent on some sort of interactive engagement activity, compared to only 8% for the non-LT class). Unfortunately, we are not able to use this data to disentangle the effect these other interactive engagement activities had on students' performances compared to the effect of the cosmology *Lecture-Tutorials* by themselves.

We repeated these observations in the Fall 2010. This time we made observations for 76% of the LT class's total class time and 85% of the non-LT class's total class time. In both cases, we observed all classes that covered cosmology. The LT class spent 18% of its total class time on cosmology and achieved an average normalized gain (across all four survey forms) of 0.31, while the non-LT class devoted 29% of its total class time to cosmology and achieved an average normalized gain of 0.27.

These observations provide insight into a commonly held belief about the effectiveness of interactive learning strategies and time on task. Specifically, faculty often state that LT students should do better than non-LT students simply because using the *Lecture-Tutorials* forces students to spend more time on a particular topic. This is not true. In both semesters, the LT classes spent a smaller percentage of their class time on cosmology and achieved larger average normalized gains than the non-LT classes. This result is even more significant when one considers the fact that the LT instructors actually gave up lecture time in order to accommodate the new cosmology *Lecture-Tutorials* into their course. This shows that while cosmology may be conceptually challenging for many students, increasing the amount of time spent lecturing about cosmology is not an effective way to help students to develop more expert-like understandings.

The benefits of interactive engagement in general and the new cosmology *Lecture-Tutorials* in particular are also demonstrated by the fact that both the LT and non-LT classes in the Fall 2010 spent a significant amount of time on a non-*Lecture-Tutorial* interactive engagement activity: think-pair-share. Specifically, 15% of the time the LT class spent teaching cosmology was used for think-pair-share (another 39% was used on the new cosmology *Lecture-Tutorials*). The non-LT class used think-pair-share for 19% of its time teaching cosmology. Research shows that think-pair-share (or Peer Instruction, as it is also commonly referred) can significantly improve students' mastery of concepts (Crouch & Mazur, 2001). The fact that the LT class in the Fall 2010 achieved a higher average normalized gain (across all four survey forms) than the non-LT class, despite the fact that they both used the research-validated instructional activity think-pair-share, supports the idea that the *Lecture-Tutorials* are effective at helping students to develop a level of conceptual mastery beyond what they can achieve from just think-pair-share.

Summary and Discussion

In this paper, we presented evidence that the new suite of five cosmology *Lecture-Tutorials* promotes a discourse among students in the traditional lecture portion of the Astro 101 classroom that improves their conceptual understandings of commonly taught cosmology topics, such as the evidence for dark matter, the expansion of the universe, and the Big Bang. We found many instances in which LT classes achieved larger normalized gains than non-LT classes. These results support previous studies (LoPresto & Murrell, 2009; Prather et al., 2004) that provide evidence to support the claim that *Lecture-Tutorials* provide an effective instructional strategy for promoting the deep intellectual engagement that facilitates student learning in the Astro 101 classroom.

That being said, we also observed several cases in which the LT students did not significantly outperform the non-LT students. For example, the LT classes and the non-LT classes had minimal normalized gains on Form A. We attribute this low overall gain among all classes (LT and non-LT) to the fact that students had to provide very detailed and sophisticated explanations using complex chains of reasoning in order to score highly on our assessment rubric for the items on Form A (Papers 2–4). Note that, as part of our research methodology, students were not given extra points (nor were they penalized) for the completeness and correctness of their responses to any of the items on any of the four surveys. When given open-response survey items and no great incentives for providing exhaustive and coherent written responses, it is not surprising that students tend to provide short, simplified, and incomplete answers to our research questions, even when they possess more sophisticated ideas in their minds. As an example, consider item 3 on Form A:

> 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.

In order to earn the maximum score, a student's response had to state that (1) the rate at which the universe is expanding must be changing over time since the slope of the graph changes with respect to distance, (2) a universe that is expanding faster and faster over time must have a slope that gets steeper over time, and (3) the slope must be flatter at large distances since we are looking further back in time as we look farther away in distance. Because the item does not explicitly prompt students to provide this level of detail, we cannot tell if students who failed to provide all these details did so because they do not know them or because they did not feel compelled to write a lengthy response. Thus, the nature of the items and our analysis methodology for Form A may have suppressed any differences that exist between the LT and non-LT students on this construct. Future studies should re-examine the effect of the new cosmology *Lecture-Tutorials* by using a different set of assessment items that explicitly probe each of the reasoning elements that students must use to correctly interpret Hubble plots.

Throughout this analysis, we have seen significant variation in the normalized gains for each survey form and significant variation in the normalized gains for individual LT and non-LT classes. For those classes (both LT and non-LT) that provided answers to every survey form ($N = 11$), we see a range of average normalized gain (for all forms) between 0.10 and 0.47, with a mean for this group of classes of 0.27 and a standard deviation of 0.10. This variation in normalized gains is important because it points to the fact that what instructors do in the classroom really matters—perhaps more than any other aspect of the course instruction examined by this study. Simply 'plugging in' the *Lecture-Tutorials* into your current class is likely not going to be sufficient to help students to develop a conceptually rich understanding of cosmology (or any other topic for that matter). Implementation matters. How one implements interactive learning activities may well be the most significant variable in explaining the variation in learning gains from class to class. This is consistent with the results from other large-scale national studies that have shown that other factors, such as institution type, class size, gender, ethnicity, prior math and science coursework, GPA, and primary language may all be of secondary importance compared to instructors' implementation practices (Hake, 1998; Prather et al., 2009; Rudolph et al., 2010).

We believe that differences in instructors' implementation practices may help to explain why the LT classes in the Fall 2010 performed lower than the non-LT classes on Forms B and C. Fall 2010 was the first semester in which instructors from outside our research group used the new cosmology *Lecture-Tutorials*. In fact, most of the instructors who volunteered in the Fall 2010 had not previously participated in our study. Since these classes were scattered across the USA (Paper 1), we could not observe each instructor's pedagogical practices in order to establish how well these instructors integrated the *Lecture-Tutorials* into the overall norms and culture of their classes. For example, we do not know if instructors provided enough time for students to work through the *Lecture-Tutorials*. We do not know if instructors simply provided answers to the *Lecture-Tutorials* (note that they were designed to be used as ungraded in-class activities, not for-credit, graded, or homework activities) or if they guided student learning groups to construct their own answers. We do not know if instructors created a classroom culture and course feedback that motivated students to work collaboratively and come to consensus on their answers when completing their *Lecture-Tutorials*. We do not know if *Lecture-Tutorials* were a frequently used and integral part of the course curriculum, or if the new cosmology *Lecture-Tutorials* were simply viewed by students as a one-time set of activities 'tacked on' to the end of the semester (note that cosmology is commonly the last topic taught in a semester of astronomy). In short, we do not know if the cosmology *Lecture-Tutorials* were implemented in a manner consistent with the best practices for this instructional strategy (Brogt, 2007; Prather et al., 2004).

Investigating how instructors implement research-proven instructional strategies into their existing classes is an important topic for future science education research studies. Research shows that while many instructors are aware of and interested in research-validated interactive engagement activities, they may exhibit significant variations in their implementations of those activities, even to the point where they disregard research-based best practices (Dancy & Henderson, 2010; Henderson & Dancy, 2009; Turpen & Finkelstein, 2009). Such implementation differences are sometimes due to the unique situations and institutional cultures in which instructors work (Henderson & Dancy, 2007), although instructors show significant variations in their implementation practices even when they are subject to these same situational constraints (Turpen & Finkelstein, 2009). Differences in implementation practices significantly influence classroom norms and students' perceptions (Turpen & Finkelstein, 2010).

What are the 'take home' messages of this study? First, the five new cosmology *Lecture-Tutorials* can elevate Astro 101 students' understandings of cosmology topics beyond what other instructional strategies can accomplish. This is significant, since up until now there has been a dearth of studies on students' conceptual and reasoning difficulties with cosmology, as well as a lack of research-validated activities that help students to overcome those difficulties. However, this elevation in student understanding is not automatic. As with many other research-validated curricular activities, the *Lecture-Tutorials* cannot simply be inserted into a course with little thought as to how they must be integrated into the larger classroom environment. Again, we assert that the variations in instructors' implementation practices are the most likely explanation for the range of learning gains we observed as part of this study. Future studies must investigate how an instructor's pedagogical content knowledge, the context of the classroom environment they create, and the specific implementation practices they employ when using the *Lecture-Tutorials* translate into student learning. Through this research, we can begin to better understand not only which combinations of these complex variables seem to be critical to student achievement, but also how best to design transformative professional development experiences for instructors who have or will be creating classrooms that use proven-interactive teaching methods.

Acknowledgements

This project would not have been possible without the contributions of many people, including members of the Center for Astronomy Education (CAE), especially Gina Brissenden and Sébastien Cormier, and the many instructors and students who graciously agreed to participate. We owe special thanks to Kim Coble, Seth Hornstein, and Tim Chambers who helped us to review and edit early versions of the *Lecture-Tutorials*. We also received valuable assistance from the Physics Education Research Group at the University of Colorado at Boulder, especially Noah Finkelstein. This material is based in part upon work supported by the National Science Foundation under grant numbers 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.

References

American Education Research Association, American Psychological Association, & the National Council on Measurement in Education (AERA, APA, & NCME) (1999). *Standards for educational and psychological testing*. Washington, DC: American Education Research Association.

Brogt, E. (2007). A theoretical background on successful implementation of Lecture-Tutorials. *Astronomy Education Review*, *6*(1), 50–58.

Clement, J., Brown, D.E., & Zeitsman, A. (1989). Not all preconceptions are misconceptions: Finding 'anchoring conceptions' for grounding instruction of students' intuitions. *International Journal of Science Education*, *11*(5), 554–565.

Crouch, C.H., & Mazur, E. (2001). Peer Instruction: Ten years of experience and results. *American Journal of Physics*, *69*(9), 970–977.

Dancy, M., & Henderson, C. (2010). Pedagogical practices and instructional change of physics faculty. *American Journal of Physics*, *78*(10), 1056–1063.

Deming, G., & Hufnagel, B. (2001). Who's taking ASTR 101? *The Physics Teacher*, *39*(6), 368–369.

Elby, A. (2001). Helping physics students learn how to learn. *American Journal of Physics*, *69*(7), S54–S64.

Erickson, F. (1986). Qualitative methods in research on teaching. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119–161). New York, NY: Macmillan Publishing Company.

Fraknoi, A. (2002). Enrollments in astronomy 101 courses: An update. *Astronomy Education Review*, *1*(1), 121–123.

Hake, R.R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, *66*(1), 64–74.

Henderson, C., & Dancy, M.H. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics: Physics Education Research*, *3*(2), 020102.

Henderson, C., & Dancy, M.H. (2009). Impact of physics education research on the teaching of introductory quantitative physics in the United States. *Physical Review Special Topics: Physics Education Research*, *5*(2), 020107.

Hudgins, D.W., Prather, E.E., Grayson, D.J., & Smits, D.P. (2006). Effectiveness of collaborative ranking tasks on students' understanding of key astronomy concepts. *Astronomy Education Review*, *38*(4), 373–392.

Kane, M.T. (1992). An argument-based approach to validity. *Psychological Bulletin*, *112*(3), 527–535.

LoPresto, M.C., & Murrell, S.R. (2009). Using the star properties concept inventory to compare instruction with lecture tutorials to traditional lectures. *Astronomy Education Review*, *8*(1), 010105.

Lyman, F.T. (1981). The responsive classroom discussion: The inclusion of all students. In A. Anderson (Ed.), *Mainstreaming digest* (pp. 109–113). College Park, MD: University of Maryland Press.

McDermott, L.C. (1991). Millikan Lecture 1990: What we teach and what is learned—Closing the gap. *American Journal of Physics*, *59*(4), 301–315.

Prather, E.E., Rudolph, A.L., Brissenden, G., & Schlingman, W.M. (2009). A national study assessing the teaching and learning of introductory astronomy. Part I. The effect of interactive instruction. *American Journal of Physics*, *77*(4), 320–330.

Prather, E.E., Slater, T.F., Adams, J.P., Bailey, J.M., Jones, L.V., & Dostal, J.A. (2004). Research on a Lecture-Tutorial approach to teaching introductory astronomy to nonscience majors. *Astronomy Education Review*, *3*(2), 122–136.

Prather, E.E., Slater, T.F., Adams, J.P., Brissenden, G., & Dostal, J.A. (2008). *Lecture-Tutorials for introductory astronomy* (2nd ed.). San Francisco, CA: Pearson Education, Inc.

Rudolph, A.L., Prather, E.E., Brissenden, G., Consiglio, D., & 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*(1), 010107.

Slater, T., Adams, J.P., Brissenden, G., & Duncan, D. (2001). What topics are taught in introductory astronomy courses? *The Physics Teacher*, *39*(1), 52–55.

Turpen, C., & Finkelstein, N.D. (2009). Not all interactive engagement is the same: Variations in physics professors' implementations of Peer Instruction. *Physical Review Special Topics: Physics Education Research*, *5*(2), 020101.

Turpen, C., & Finkelstein, N.D. (2010). The construction of different classroom norms during Peer Instruction: Students perceive differences. *Physical Review Special Topics: Physics Education Research*, *6*(2), 020123.

Wallace, C.S., Prather, E.E., & Duncan, D.K. (2011a). A study of general education astronomy students' understandings of cosmology. Part I. Development and validation of four conceptual cosmology surveys. *Astronomy Education Review*, *10*(1), 010106.

Wallace, C.S., Prather, E.E., & Duncan, D.K. (2011b). 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*(1), 010107.

Wallace, C.S., Prather, E.E., & Duncan, D.K. (2012a). 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*, *11*(1), 010103.

Wallace, C.S., Prather, E.E., & Duncan, D.K. (2012b). A study of general education astronomy students' understandings of cosmology. Part IV. Common difficulties students experience with cosmology. *Astronomy Education Review*, *11*(1), 010104.