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

Astronomy Education Review

Volume 6, Aug 2007 - Mar 2008
Issue 2

First Results from the Light and Spectroscopy Concept Inventory

by **Erin M. Bardar**
Boston University and TERC
Received: 09/30/07, Revised: 11/19/07, Posted: 01/24/08

The Astronomy Education Review, Issue 2, Volume 6:75-84, 2008

Abstract

This article presents results from a two-semester field test of the Light and Spectroscopy Concept Inventory (LSCI). Statistical analysis indicates that the LSCI has the sensitivity to measure statistically significant changes in students' understanding of light-related topics due to instruction in introductory astronomy courses and to distinguish the relative effectiveness of traditional (primarily lecture) and active engagement treatments in introductory college astronomy courses.

1. INTRODUCTION

The development and validation of the Light and Spectroscopy Concept Inventory (LSCI) are described in a separate article (Bardar et al. 2007). In this article, I present results from a two-semester nationwide field test of the LSCI and describe the analyses performed to determine whether the inventory has the sensitivity to (1) measure a statistically significant change in student understanding of topics related to light and spectroscopy due to the instruction provided in a semester-long introductory college astronomy course, and (2) differentiate between traditional instruction (primarily lecture) and active engagement treatments.

The article is organized as follows. In Section 2, I describe the statistical techniques used to analyze the data. In Section 3, I present the results from the two-semester field test. In Section 4, I discuss the implications of the results and suggest further work.

2. ANALYSIS TECHNIQUES

Testing took place during the 2005–2006 academic year with student examinees enrolled in 34 course sections at 26 colleges and universities. The LSCI was administered during the first week of classes to establish a baseline of content knowledge, and then readministered at the end of the course to determine the effectiveness of instruction in promoting conceptual gains.

Statistical *t*-tests were performed to compare pretest and posttest means for each course section in order to determine if the score changes were statistically significant. This analysis technique was also used to explore whether the LSCI was able to differentiate between traditional (primarily lecture) and active engagement (learner-centered) treatments. In addition to *t*-tests for statistical significance, effect sizes were calculated for each course section to quantify "practical significance"—the amount of shift in conceptual understanding resulting from the instructional intervention employed.

2.1 *T*-tests

A *t*-test is a hypothesis test used to determine if the difference between two means is statistically significant or due to random sampling fluctuations. The null hypothesis for this analysis states that the preinstruction and postinstruction means for a course section are equal (i.e., instruction produced no significant change in conceptual understanding). The alternative hypothesis, that the postinstruction mean differs by a statistically significant amount (higher or lower), is only tenable if sufficient evidence is provided to reject the null hypothesis. The *t*-test assumes that the population under inspection represents a normal distribution. A departure from normality in the distribution may lead to an invalid dismissal or retention of the null hypothesis. However, if the score distribution varies nonnormally, the test can still produce accurate results when applied to a sufficiently large sample, generally agreed to be ≥ 30 (Sprinthall 2003).

SPSS (Statistical Package for the Social Sciences) 12.0 for Windows was used to perform *t*-tests on all LSCI data. The *t*-test procedure generates a statistical value, *t*, and an associated probability level, *p*. The value of *t* is calculated by the following formula:

$$t = \frac{\text{sample mean} - \text{hypothetical mean of the null hypothesis}}{\text{standard deviation of the sample mean} \div \text{square root of the sample size}}$$

The significance level, *alpha*, of a *t*-test represents the maximum likelihood of rejecting a true null hypothesis. For a 95% confidence interval, the significance level is *alpha* = 0.05. If the probability value *p* is less than or equal to *alpha*, there is a statistically significant difference between the means of the two compared populations, and the null hypothesis can be rejected. If *p* > *alpha*, the null hypothesis must be retained.

A sample SPSS *t*-test output is shown in Figure 1. In general, the larger the value of *t*, the larger the difference between the two means being compared. If the confidence interval contains zero, the null hypothesis cannot be rejected. The column heading *df* represents the number of degrees of freedom, equal to one less than the sample size (*N*-1). The *p*-value is reported in the table as "Sig. (2-tailed)." This indicates that the alternative hypothesis is nondirectional—it was chosen to state the alternative hypothesis such that the effect of instruction on students' understanding of the concepts addressed by the LSCI could be either negative or positive. If the alternative hypothesis was altered to presuppose a statistically significant gain only, the value of *p* should be halved.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Post-Test (AE)	688	11.96	4.852	.185

One-Sample Test

	Test Value = 6.16					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Post-Test (AE)	31.351	687	.000	5.799	5.44	6.16

Figure 1. Sample SPSS *t*-test output tables. The top table shows the sample size, mean posttest score, standard deviation, and standard error of the mean for the test sample (postinstruction). The bottom table displays the results of a single-sample *t*- test comparing the posttest sample mean (11.96) with the pretest mean (6.16). Because the null hypothesis would be that the means are the same, the hypothetical mean of the null hypothesis (the Test Value) is equal to the pretest mean (6.16). Other values displayed in the output table are the calculated *t* value; the degrees of freedom (*df*); the associated probability value for incorrectly rejecting the null hypothesis (Sig. 2-tailed); the difference between the test sample mean score and the Test Value (Mean Difference); and the upper and lower boundaries of the 95% confidence interval.

2.2 Effect Size

Effect Size (*ES*), also known as Cohen's *d* (Cohen 1988), is a measure of the magnitude of the effect created by an instructional intervention or treatment and can be interpreted as the percent of nonoverlap of the treated group's scores with those of the group before or without instruction. Mathematically, *ES* relates the ratio of the difference in means from two groups (here taken to be preinstruction and postinstruction) to the pooled (average) standard deviation of the two means. *ES* does not specify the number of points by which two group means differ, but rather is expressed in units of standard deviation. Therefore, its meaning is the same within a single study or across multiple studies, and *ES* is said to have "practical significance" (Kirk 1999). An effect size of 0.2 is considered small, 0.5 is considered medium, and 0.8 or greater is considered to be a large effect (Cohen 1988).

3. RESULTS

During the first semester of the LSCI field test, results were obtained for 548 students prior to instruction and 368 students after instruction. Class sizes ranged from 7 to 68 students for the 14 course sections at 11 colleges and universities participating in both halves of the study. The second semester sample size was considerably larger, with a preinstruction *N* of 1,518 and a postinstruction *N* of 1,377, representing students enrolled in 20 course sections from 15 colleges and universities. Class enrollments ranged in size from 10 to 154 students.

3.1 General Results

Preinstruction means for individual course sections ranged from 21.8% to 65.4%, and postinstruction means ranged from 23.5% to 76.2%. The higher of the values for each range of scores came from section C5. Whereas all other courses included in the study were on-campus undergraduate courses for nonmajors at four-year colleges and universities, section C5 was an online course for in-service teachers. A *t*-test confirmed that the scores from this group should be considered as a separate student population. Excluding this course, preinstruction means ranged from 21.8% to 34.7%, and postinstruction means ranged from 23.5% to 52%. However, because baseline scores were obtained during the first week of instruction, *t*-test results and effect size have the same meaning for section C5 as they do for all other sections included in the study.

T-test results showed that the improvement in scores for 28 of the 34 course sections was statistically significant at the 95% confidence level. Of the six sections in which the measured change in scores due to the instructional intervention was not found to be statistically significant, there were likely several contributing factors. These include small class size, lack of complete material coverage (as reported by instructors), and insufficient depth-of-coverage of material (also as reported by instructors).

Medium to large effect sizes were calculated for all but one (C9) of the course sections exhibiting statistically significant improvement. For the courses exhibiting no statistically significant improvement due to instruction, effect size or practical significance was accordingly unsatisfactory. The instructional interventions for four of these sections produced small to medium magnitude positive effects, and two produced small or medium negative effects.

3.2 Traditional Instruction versus Active Engagement Techniques

For the purpose of this study, a course considered to employ "active engagement treatments" must have dedicated a significant portion of instructional time to incorporating student-centered activities such as peer instruction (Green 2003; Mazur 1997), Lecture-Tutorials (Prather et al. 2004), or homelabs (see Bardar 2006; Bardar & Brecher 2008). Instructors participating in the LSCI field test were asked to complete a brief survey describing the nature of both the lecture and laboratory/discussion (if applicable) components of his or her course. These surveys were used to determine the relative amounts of active engagement techniques employed in each of the course sections. Because the amount of active engagement incorporated into each course was self-reported by instructors, there may be some inaccuracies in the designations of certain courses as active engagement (AE) or traditional (T). However, because there was not an overabundance of instructors reporting heavy use of active engagement techniques, it is likely that instructors were as honest as possible in the assessment of their courses.

In total, 10 sections (B0, B4, B7, B8, B9, C0, C4, C8, D1, and D5) reported extensive use of peer instruction (PI) and Lecture-Tutorials (LT) as part of the lecture component of their courses. *T*-tests comparing the mean scores of these active engagement courses with both the full sample mean and with the mean of all traditional lecture-based courses revealed statistically significant differences.

Section A0 did not employ peer instruction or tutorial techniques during regular class meetings but made extensive use of interactive hands-on, inquiry-based light and optics homelabs (Bardar 2006; Bardar & Brecher 2008) as a major portion of the course. Section A0 was therefore considered to be an active engagement (AE) course. A *t*-test was performed to determine if the postcourse mean should indeed be considered to have come from the same population of scores as the other interactive courses. This test showed that the null hypothesis could not be rejected ($t = 0.519$, $p = 0.607 > alpha = 0.05$), and the conclusion was drawn that A0 was accurately included in the active engagement category.

Another *t*-test comparing all 11 active engagement courses with the more traditional (T) lecture-centered courses yielded a *t*-value of 9.792 (p =.000), indicating that the LSCI was successfully able to determine a statistically significant difference between traditional instruction and active engagement treatments. Table 1 summarizes the full set of field test results.

Table 1. Summary of LSCI Field Test Results

Section	***N* (pre)**	**Pre %**	***N* (post)**	**Post %**	**Gain %**	**<g>**	***t***	***p***	**Effect Size**	**AE/T**
A0	68	32.3	38	51.8	19.5	0.29	6.52	0.000	1.31	AE (H)
A1	42	23.9	37	32.6	8.7	0.11	4.37	0.000	0.79	T
A2	40	25.7	28	35.4	9.7	0.13	5.17	0.000	1.08	T
A3	20	26.3	16	24.8	-1.5	-0.02	-0.14	0.891	-0.14	T
A4	33	24.0	31	32.6	8.6	0.11	3.61	0.001	0.76	T
A5	31	25.6	18	31.4	5.8	0.08	1.48	0.157	0.45	T
A6	20	27.9	20	23.5	-4.4	-0.06	-1.95	0.066	-0.42	T
A7	10	23.5	8	28.4	4.9	0.06	1.07	0.321	0.47	T
B0	46	24.9	39	50.5	25.6	0.34	10.35	0.000	2.01	AE (PI)
B1	31	22.6	25	39.2	16.6	0.21	4.83	0.000	1.28	T

B2	31	25.7	27	39.3	13.6	0.18	4.06	0.000	1.03	T
B3	33	21.8	30	37.9	16.1	0.21	4.56	0.000	1.12	T
B4	53	29.9	34	50.0	20.1	0.29	5.25	0.000	1.13	AE (PI)
B5	24	22.9	17	33.3	10.4	0.13	3.72	0.002	1.03	T
B6	49	29.7	35	34.4	4.7	0.19	1.82	0.077	0.34	T
B7	136	22.9	105	51.2	28.3	0.37	16.73	0.000	2.04	AE (LT)
B8	129	23.0	107	52.0	29.0	0.38	14.97	0.000	2.00	AE (LT)
B9	94	23.3	64	35.5	12.2	0.16	5.75	0.000	0.93	AE (LT)
C0	37	25.1	35	37.7	12.6	0.17	5.65	0.000	1.16	AE (LT)
C1	28	25.1	15	38.5	13.4	0.18	4.06	0.001	1.27	T
C2	37	23.9	31	33.1	9.2	0.12	3.62	0.001	0.68	T
C3	35	25.2	26	28.6	3.4	0.05	1.71	0.099	0.33	T
C4	32	22.0	26	28.8	6.8	0.09	3.25	0.003	0.74	AE (PI)
C5	13	65.4	10	76.2	10.8	0.31	2.30	0.047	0.65	online
C6	43	34.7	42	42.9	8.2	0.13	3.06	0.004	0.49	T
C7	154	22.6	140	27.8	5.2	0.07	5.63	0.000	0.50	T
C8	100*	22.5	67	34.1	11.6	0.15	5.37	0.000	0.93	AE (LT)
C9	145	22.6	169	26.1	3.5	0.05	3.70	0.000	0.33	T

D0	149	23.1	143	38.8	15.7	0.20	12.62	0.000	1.30	T
D1	146	22.1	147	50.4	28.3	0.36	22.35	0.000	2.18	AE (PI)
D2	124	28.9	85	50.4	21.5	0.30	12.34	0.000	1.52	T
D3	13	24.6	15	35.9	11.3	0.15	2.99	0.012	0.83	T
D4	50	28.9	86	42.1	13.2	0.19	7.48	0.000	0.91	T
D5	100*	22.5	24	39.6	17.1	0.22	4.99	0.000	1.66	AE (LT)

Note: Active engagement courses are labeled AE, with the specific type of technique in parentheses. H = homelabs; PI = peer instruction; LT = Lecture-Tutorials. Traditional courses are labeled T. "Online" denotes the online course for practicing teachers, which does not fall within either the AE or T designation. The * next to the *N* (pre) values for sections C8 and D5 indicates that these sections were combined for the pretest with a total *N* of 100 but were treated separately for posttest analysis.

Figure 2 shows an alternate representation of the field test results, through a plot of gain score versus pretest score for each of the participating course sections. Normalized gain (<g>) is defined as the ratio of the actual gain to the maximum possible gain. The value of <g> can be interpreted as the percentage of possible gain achieved, that is, <g> = 0.5 means that 50% of the maximum increase in score possible was attained. The average normalized gain score for courses employing active engagement techniques, $<<g>>_{AE} = 0.255 \pm 0.3$, was nearly twice that measured for traditional lecture courses, $<<g>>_{T} = 0.129 \pm .09$.

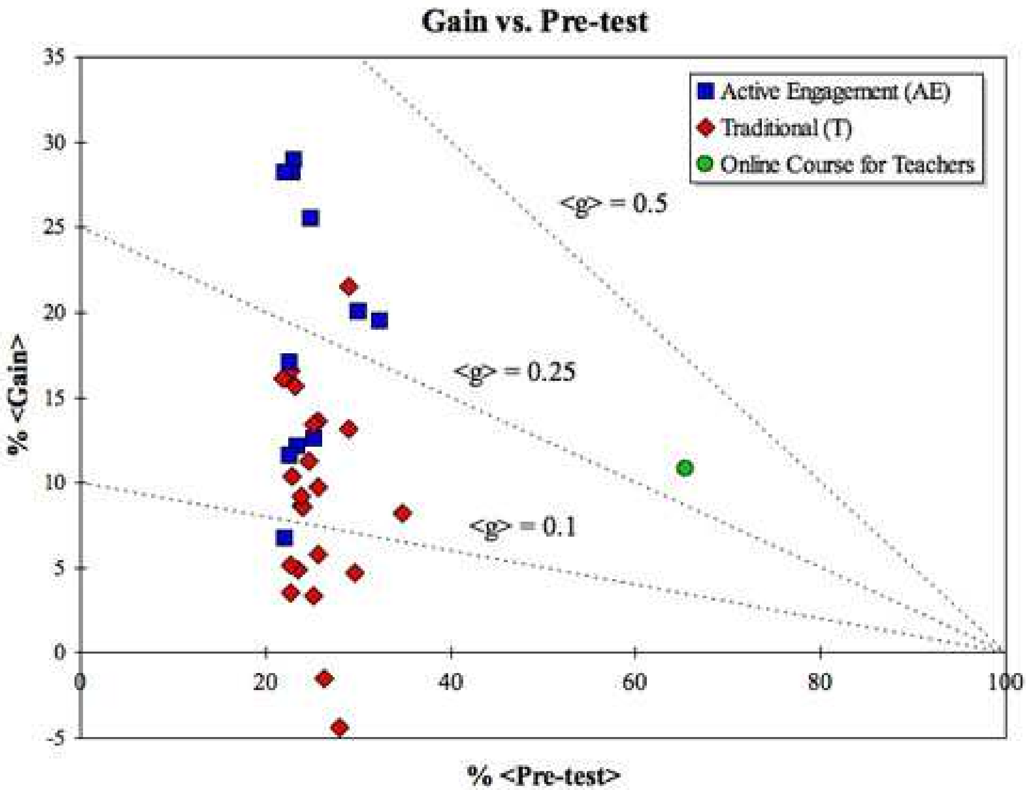

Figure 2. Comparison of the relative effectiveness of instructional interventions as measured by the LSCI. <g> indicates the normalized gain index. <<g>> indicates the average normalized gain. The dotted lines indicate constant values of normalized gain, <g>, of 0.1, 0.25, and 0.5.

4. DISCUSSION AND SUGGESTIONS FOR FUTURE WORK

Statistically significant gains in student understanding were measured for 28 of the 34 sections that participated in both preinstruction and postinstruction testing. Tests of statistical significance were also able to successfully differentiate courses employing learner-centered techniques from courses relying heavily on lecture as the primary mode of instruction. Although the course section employing homelabs retained lecture as the primary in-class method of instruction, the introduction of take-home learner-centered materials and activities proved to be as effective as, or more effective than, peer instruction/tutorial methods.

Small number statistics and insufficient coverage of topics on the inventory contributed to the lack of statistically significant improvements in six of the course sections participating in the field test. Although topics addressed by the inventory were thoroughly researched to be applicable to the widest possible range of introductory college astronomy courses, there is no standard curriculum at the college/university level as there is at the elementary and high school levels. For this reason, the range and depth of topics covered are at the discretion of the instructor or his or her institution and will necessarily vary to some degree.

With the exception of the online course for in-service teachers (C5), no section achieved a mean postinstruction score above 52%. The reality that students are performing at or below the 50% level after instruction on a test comprising material that is central to the introductory astronomy course indicates that the community needs to purposefully examine its instructional practices and find a way to improve the teaching and learning in the classroom.

With the LSCI, instructors will be able to assess the effectiveness of their instruction in promoting conceptual change with regard to light-related topics. This has practical applications for instructors who are required to provide their institution with evaluation of their courses. Wide-scale use of the LSCI can provide the larger community of astronomy instructors with the evidence needed to enact widespread reform in the way that the introductory college astronomy survey course is taught by demonstrating which instructional techniques are most successful in producing meaningful learning gains.

To verify the results of this study, more data should be collected. This expanded study should include more active engagement courses, as well as clusters of courses implementing the same instructional technique, to provide more robust comparison data. Additionally, courses employing multiple active engagement techniques should be targeted for participation in the study. Future studies with LSCI data might also aim to investigate whether the instrument is sensitive to gender bias or differences between high- and low-achieving students (as determined by overall course performance or grade).

Acknowledgment

Portions of this article are contained in Erin Bardar’s PhD thesis. For a copy of the full dissertation, please send a request to Erin at ebardar@gmail.com.

Funding for this project was provided in part by NSF Grant No. DUE-0125992 and NASA GSRP Grant No. NGT5-60482.

I would like to acknowledge Ed Prather, Tim Slater, and Ken Brecher for their contributions to the development of the LSCI. I would also like to extend a sincere thank you to the instructors and students who participated in the LSCI field test.

References

Bardar, E. M. 2006, "Development and Analysis of Spectroscopic Learning Tools and the Light and Spectroscopy Concept Inventory for Introductory College Astronomy," Unpublished doctoral dissertation, Boston University, Boston, MA.

Bardar, E. M., & Brecher, K. 2008, "Project LITE Educational Materials and Their Effectiveness as Measured by the Light and Spectroscopy Concept Inventory," *Astronomy Education Review*, 6(2). http://aer.noao.edu/cgi-bin/article.pl?id=255.

Bardar, E. M., Prather, E. E., Brecher, K., & Slater, T. F. 2007, "Development and Validation of the Light and Spectroscopy Concept Inventory," *Astronomy Education Review*, 5(2), 103. http://aer.noao.edu/cgi-bin/article.pl?id=225.

Cohen, J. 1988, *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.), Hillsdale, NJ: Lawrence Erlbaum.

Green, P. J. 2003, *Peer Instruction for Astronomy*, Upper Saddle River, NJ: Prentice Hall.

Kirk, R. E. 1999, *Statistics: An Introduction* (4th ed.), Belmont, CA: Wadsworth/Thomson Learning.

Mazur, E. 1997, *Peer Instruction*, Upper Saddle River, NJ: Prentice Hall.

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 for Non-Science Majors," *Astronomy Education Review*, 3(2), 122. http://aer.noao.edu/cgi-bin/article.pl?id=132.

Sprinthall, R. C. 2003, *Basic statistical analysis* (7th ed.), Boston: Allyn & Bacon.

ÆR
75 - 84