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

Astronomy Education Review

Volume 3, Oct 2004 - Apr 2005
Issue 2

Assessment of Teaching Approaches in an Introductory Astronomy College Classroom

by
Received: 08/12/04, Revised: 02/16/05, Posted: 03/11/05

The Astronomy Education Review, Issue 2, Volume 3:178-186, 2005

Abstract

In recent years, there have been calls from the astronomy education research community for the increased use of learner-centered approaches to teaching, and systematic assessments of various teaching approaches using such tools as the Astronomy Diagnostic Test 2.0 (ADT 2.0). The research presented is a response to both calls. The ADT 2.0 was used in a modified form to obtain baseline assessments of introductory college astronomy classes that were taught in a traditional, mostly didactic manner. The ADT 2.0 (modified) was administered both before and after the completion of the courses. The courses were then altered to make modest use of learner-centered lecture tutorials. The ADT 2.0 (modified) was again administered before and after completion of the modified courses. Overall, the modest learner-centered approach showed mixed statistical results, with an increase in effect size (from medium to large), but no change in normalized gain index (both were low). Additionally, a mathematically rigorous approach showed no statistically significant improvements in conceptual understanding compared with a mathematically nonrigorous approach. This study will interpret the results from a variety of perspectives. The overall implementation of the lecture tutorials and their implications for teaching will also be discussed.

1. INTRODUCTION

There is a growing call by the astronomy education community for more assessment as a tool for improving astronomy education (Brissenden, Slater, & Mathieu 2002). This assessment can be for course content determination, astronomy misconception inventory, or development of instructional methods. The Astronomy Diagnostic Test 2.0 has been developed over the past several years to assess conceptual understanding in astronomy (Zeilik 2003; Hufnagel 2002). Other calls by the astronomy education community (Straits & Wilke 2003), as well as the science education community as a whole (Walczyk & Ramsey 2003), have included the greater use of learner-centered instructional approaches in introductory college-level science classrooms. Implementing a learner-centered approach can be a difficult task. For example, confusion can arise in the course, contributing to a significant reduction in students' self-efficacy with regard to science (Straits & Wilke). Perhaps because of these difficulties, didactic, instructor-centered teaching methods are still alive and well in college science classrooms (Walczyk & Ramsey). More assessment of learner-centered approaches to teaching introductory college astronomy, especially in my classes, seems to be an area worth exploring.

This study was conducted to evaluate the differences in astronomy conceptual understanding, as measured by the ADT 2.0, when a modest learner-centered approach is compared with a didactic, instructor-centered approach. The differences in conceptual understanding for a mathematically rigorous course compared with a mathematically nonrigorous course will also be evaluated. This research was conducted in the tradition of "action research" (Gall, Gall, & Borg 2003), although this work can be viewed in light of the larger body of work to generalize to other situations.

2. COURSE DESCRIPTIONS

PHYS 120 (The Solar System) is intended to be a survey course in astronomy that is accessible to a wide range of students and is considered mathematically nonrigorous. It uses minimal mathematics and emphasizes conceptual understanding of astronomy. PHYS 220 (General Astronomy I), on the other hand, is designed to be a more challenging course and is considered mathematically rigorous. It is a problem-solving course and makes much more extensive use of mathematics, in addition to a parallel goal of conceptual understanding of astronomy. Both courses cover the basic physics concepts required to understand astronomy. Each of these courses was taught in a planetarium, so observational astronomy was a significant part of both courses. The later part of each course covers the Solar System and formation of planetary systems. Stellar evolution, galactic astronomy, and cosmology are not covered in these courses. PHYS 220 fulfills a specific science requirement in the general education program of James Madison University (JMU), while PHYS 120 can only serve as an elective. The two classes generally have a similar student makeup, with five to eight science majors and class sizes from 30 to 45 students.

I compared each item of the ADT 2.0 with the content of both PHYS 120 and PHYS 220, and found that only one question was not specifically covered in these courses. Question 16 from the ADT 2.0 dealt with cosmology and was not counted toward the final score. This was the modified version of the ADT 2.0 that was used in this study.

3. EXPERIMENTAL DESIGN

I taught both PHYS 220 and PHYS 120 in the fall of 2002. The courses were delivered in a didactic fashion, with minimal inquiry-based discussions throughout each course. We did not attempt learner-based instruction. The ADT 2.0 (modified) was administered to each class as both a pretest and a posttest to serve as a baseline for future comparison. The two courses differed in their levels of mathematical rigor, as described in the previous section. The makeup of each class was generally the same, with roughly equal numbers of science and non–science majors in each class. Even though there was not a random assignment of students into the classes, there was no reason to think that the students in the two classes differed substantially in their abilities.

In the fall of 2003, I again taught PHYS 220 and PHYS 120. The difference was that the courses made modest use of learner-centered approaches to teaching. The ADT 2.0 (modified) was administered to each class as both a pretest and a posttest in order to compare differences in conceptual understanding of astronomy with the baseline data taken in the previous year. The two courses still differed in their levels of mathematical rigor. The makeup of each class was generally the same, with roughly equal numbers of science and non–science majors in each class. There were also roughly equal numbers of male and females in each course. Even though there was not a random assignment of students into the classes, again there was no reason to think that the two classes were substantially different in their abilities. This assertion will be tested later in the study.

3.1 Inclusion in This Study

I chose to include only students with valid pretest and posttest scores, so this study used repeated measures. This eliminates the bias that could arise by low-performing students opting out of taking the posttest ADT 2.0 (modified). If a student completed only a pretest or only a posttest, his or her single completed test score was not used.

3.2 Treatment

I was both the instructor and the researcher for this study. When I began this study in 2002, I was in my third year of teaching astronomy full time. The didactic lecture mode was used in 2002. In 2003, my fourth year of teaching astronomy, I incorporated a modest amount of learner-centered lecture tutorials into my classes.

Student-centered instructional tools have become available in recent years to help facilitate learner-centered learning. These include *Interactive Lesson Guide for Astronomy 2nd Edition* (Zeilik 2001) and *Lecture-Tutorials for Introductory Astronomy* (Adams, Prather, & Slater 2003). In choosing an approach, I felt that it would be important to use a tool flexible enough to be expanded on or pared down, depending on the dynamics of the classroom. Without this flexibility, students may perceive a lack of structure in the course and not know how to focus their efforts (Straits & Wilke 2003). I found that many lessons in the *Interactive Lesson Guide for Astronomy 2nd Edition* were quite involved and required a fair amount of didactic instruction just to have the students complete the assignment.

As a contrast, *Lecture-Tutorials for Introductory Astronomy* was designed so that sections could be deleted or added depending on the dynamics of the class. Ten exercises were chosen from *Lecture-Tutorials for Introductory Astronomy* to be used in both PHYS 120 and PHYS 220 during the fall of 2003. Roughly one of these exercises was covered per week, requiring approximately 30 minutes each. These 30-minute exercises accounted for roughly one third of the 90 contact minutes per week. I have categorized this as a *modest* use of learner-centered techniques. They were not graded assignments because they were used purely as a way to engage students during class, with the aim of increasing their conceptual understanding of astronomy over and above didactic lecture methods. Students worked in groups of two to three and compared their results with those of other groups at the completion of each section. The groups were self-assembled and were determined by the normal seating configuration of the classroom. The lecture-tutorial sections from Adams, Prather, & Slater (2003) that were used are:

- (celestial) Motion (pages 3–4)
- Seasonal Stars (pages 7–9)
- The Cause of Moon Phases (pages 23–27)
- Predicting Moon Phases (page 28)
- Blackbody Radiation (pages 33–34)

- Types of Spectra (pages 37–38)
- Telescopes and Earth’s Atmosphere (pages 43–44)
- Understanding Retrograde Motion (pages 47–48)
- Orbital Period and Orbital Distance (pages 49–51)
- Temperature and Formation of Our Solar System (pages 57–58)

Care was taken so that the only significant teaching difference between 2002 and 2003 was that a learner-centered approach was used in 2003. This is important because any differences in performance between the two years could be attributed to the treatment. The same textbooks were used from year to year, and the instructor was the same.

The demographics from the ADT 2.0 (modified) were also considered to establish a level of equivalence between the years. The demographics, however, were not used to establish causal relationships for ADT outcomes. During both years, over 95% of the students were Caucasian and traditional college age (18–22). In 2002, 28.2 % of the students were science majors, compared with 23.3% in 2003. In 2002, the female/male ratio was 41:59, and in 2003 was 60:40. There was nothing directly suggesting that the two years were significantly different.

4. STATISTICAL ANALYSIS PROCEDURES

I used standard statistical analysis techniques for interpreting ADT data as outlined in Zeilik, Bisard, & Lee (2002) and Zeilik (2003). These analyses consisted of the calculation of simple descriptive statistics, normalized gain index <*g*>, and effect size. The normalized gain index is calculated by dividing the difference in posttest score and pretest score by the difference between 100% and the pretest score. So, for normalized gain index: <*g*> = (% Post - %Pre)/(100% - % Pre). The higher the normalized gain index, the greater the gain between pretest and posttest scores (Hake 1998). A normalized gain index normally ranges from 0 (lowest gain) to 1 (highest gain). Negative values are possible for <*g*>, which represent a negative gain. The normalized gain index accounts for differences in pretest scores between groups and can mitigate potential initial differences between groups. Interpretation of <*g*> values appears in section 5.2. The effect size (*ES*), or Cohen’s d, is also a measure of gain from pretest to posttest, but takes into account the standard deviations of both mean scores. The effect size is a way to quantify the difference between means of two groups. The larger the effect size, the larger this difference. It is calculated by dividing the difference between the pretest and posttest means by the standard deviation of the pretest mean. So, for effect size, *ES* = (% Post - %Pre)/(mean *SD* of Pre and Post). Interpretation of *ES* appears in section 5.2. I also used *t* tests to check for statistically significant differences between grouped means. The statistical package SPSS 12.0 for Windows was used to perform calculations.

5. RESULTS

In general, the pretest and posttest results for PHYS 120 and PHYS 220 followed the national results of the ADT 2.0, with the pretest approximately 32.4% and the posttest approximately 47.3% (Deming 2002). Comparison with national results was not the goal for this project even though such comparisons are interesting and could be performed in the future.

5.1 Results from Mathematically Rigorous Versus Mathematically Nonrigorous

Table 1 shows results obtained using combined data from 2002 and 2003. Results for *t* tests were generated by comparing data from PHYS 120 and PHYS 220. All indications point to no statistically significant differences between PHYS 120 and PHYS 220 (at the alpha = 0.05 level) when comparing pretest or posttest results, so all of the paired means for class comparisons are statistically the same.

Table 1. Comparison of ADT Results for Mathematically Rigorous (PHYS 220) and Mathematically Nonrigorous (PHYS 120) Astronomy Courses

Year	Measure	Class	Mean	*SD*	*N*	*t*	*df*	*sig*
2002 (didactic)	Pretest	120	45.00	19.15	16			
	Pretest	220	41.74	17.88	23	0.54	37	0.59
	Posttest	120	54.06	20.02	16			
	Posttest	220	54.35	17.98	23	0.05	37	0.96
	Pretest (combined)		43.08	18.23	39			
	Posttest (combined)		54.23	18.59	39	2.67	76	<0.01*
2003 (modest learner-centered)	Pretest	120	35.74	13.64	27			
	Pretest	220	33.18	17.27	33	0.63	58	0.53
	Posttest	120	55.00	14.41	27			
	Posttest	220	46.41	20.43	33	1.88	58	0.06
	Pretest (combined)		34.33	15.66	60			
	Posttest (combined)		50.17	18.36	60	5.08	118	<0.01*
2002	Pretest (combined)		43.08	18.23	39			
2003	Pretest (combined)		34.33	15.66	60	2.55	97	0.01*

* Indicates grouping is statistically different at the alpha = 0.05 level.

t = the *t* test value that measures the difference between two means. In general, the higher the *t* value, the more the two means differ.

df = degrees of freedom (the number of values that are free to vary), and is based on *N*. Because two means have been determined in each grouping already, *df* = Total *N* -2 for each grouping.

sig = the significance (or *p* level), the probability that we would determine the compared means to be the same when they are actually different. With an alpha level = 0.05, it means that a criterion has been set so that there will be 5% or less chance that the compared means will be considered the same (by chance variation) when they are actually different. Because the *t* tests are two tailed, there isn't a preconceived notion of which mean is larger.

Having no statistical differences between PHYS 120 and PHYS 220 classes indicates that a mathematically rigorous approach to teaching an introductory astronomy class does not improve conceptual understanding of astronomical topics as covered in the ADT 2.0 (modified). This does not imply that there is no place for a mathematically rigorous approach in introductory astronomy; such classes may be necessary at universities with major or minor programs in astronomy. For the purposes of this study, it means that the data from the PHYS 120 and PHYS 220 classes can be combined for comparison between years. The pretest and posttest scores are statistically different at the alpha = 0.05 level, as would be expected with gains in the posttest score. A more surprising result was that the 2002 and 2003 pretests were statistically different at the alpha = 0.05 level. Initial differences for the two groups will be mitigated with the use of the normalized gain index $\langle g \rangle$.

5.2 Results: Didactic Versus Learner Centered

The following results were obtained using data by combining PHYS 120 and PHYS 220 for each year. The normalized gain index $\langle g \rangle$ results were categorized as low for both approaches (Hake 1998). The didactic approach (2002) resulted in $\langle g \rangle = 0.20$, and the modest learner-centered approach resulted in $\langle g \rangle = 0.24$. There appear to be significant differences between 2002 and 2003 when comparing effect size. In 2002, the effect size was 0.61, which can be categorized as medium (Fan 2001). In 2003, there was an effect size of 0.93, which can be categorized as large (Fan) and exceptional (Zeilik 2003).

Table 2. Comparison of ADT Results for Didactic (Lecture-Based) and Learner-Centered Astronomy Courses

Grouping	Measure	Mean	*SD*	N	<*g*>	ES
2002 (didactic)	Pretest (combined)	43.08	18.23	39		
2002 (didactic)	Posttest (combined)	54.23	18.59	39	0.20	0.61
2003 (modest learner centered)	Pretest (combined)	34.33	15.66	60		
2003 (modest learner centered)	Posttest (combined)	50.17	18.36	60	0.24	0.93

Normalized gain index <*g*> and effect size (*ES*) results when comparing a didactic-based class (2002) with a learner-centered class using lecture tutorials (2003). The normalized gain index showed no substantial change between years, but the effect size showed a more significant difference.

5.3 Key Results

- There were no statistically significant differences between the PHYS 120 (conceptual) class and the PHYS 220 (mathematically rigorous) class at an alpha = 0.05 level.
- The normalized gain index <*g*> shows that both approaches resulted in a low gain.
- The effect size for the didactic approach was medium, while the modest learner-centered approach resulted in a large effect size.

6. IMPLICATIONS FOR TEACHING

The finding of no statistical difference between the mathematically rigorous class and the mathematically nonrigorous class was not too much of a surprise because problem solving in astronomy does not directly address conceptual understanding. From my perspective, that a student can get the correct answer for a mathematically based astronomy class does not seem to improve his or her conceptual understanding of the problem.

Even with mixed statistical results showing improvements with the learner-centered approach to teaching astronomy, this approach is a very positive step forward. It is important for students to actively participate in the learning process. Even when students resist changing into an active learning mode, it is the role of the instructor to motivate his or her students to fully participate. Only when students are fully engaged will they reach a deeper conceptual understanding of astronomy. It is likely that many of the benefits of using a learner-centered approach are intangible at this time and not fully measured by the ADT 2.0. The difficulty in developing a reliable and valid conceptual astronomy diagnostic test is echoed by others (Zeilik 2003). I agree with the teaching implications reached by Straits & Wilke (2003): "active learning strategies. . . will improve student characteristics and increase achievement." I see that the challenge is how to evaluate these student characteristics and achievement.

7. FUTURE STUDIES

The most immediate need is to continue this research to increase the sample size and reduce the possibility of nonequivalent groups. This current data set also includes demographic information, so future research questions could include gender bias, ethnic bias, attitudes toward science, and a full comparison of this current data with the National ADT 2.0 study. An increasing body of work is using the ADT 2.0 as an instrument, so meta-analysis might soon be needed to identify trends in this set of statistical studies. Future studies could also look into currently intangible benefits of using a learner-centered approach to teaching astronomy that are not currently measured by the ADT 2.0.

References

Adams, J. P., Prather, E. E., & Slater, T. F. 2003, *Lecture-Tutorials for Introductory Astronomy*, Upper Saddle River, NJ: Prentice Hall.

Brissenden, G., Slater, T. F., & Mathieu, R. D. 2002, The Role of Assessment in the Development of the College Introductory Astronomy Course, *Astronomy Education Review*, 1(1), 1.

Deming, G. L. 2002, Results from the Astronomy Diagnostic Test National Project, *Astronomy Educational Review*, 1(1), 52.

Fan, X. 2001, Statistical Significance and Effect Size in Educational Research: Two Sides of a Coin, *Journal of Educational Research*, 94(5), 275.

Gall, M. D., Gall, J. P., & Borg, W. R. 2003, *Educational Research: An Introduction*, 7th Ed. Boston: Allyn & Bacon.

Hake, R. R. 1998, Interaction-Encouragement Versus Traditional Methods: A Six-Thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses, *American Journal of Physics*, 66(1), 64.

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

Straits, W. J., & Wilke, R. R. 2003, Activities-Based Astronomy: An Evaluation of an Instructor's First Attempt and Its Impact on Student Characteristics, *Astronomy Education Review*, 2(1), 46.

Walczyk, J. J., & Ramsey, L. L. 2003, Use of Learner-Centered Instruction in College Science and Mathematics Classrooms, *Journal of Research in Science Teaching*, 40(6), 566.

Zeilik, M. 2001, *Interactive Lesson Guide for Astronomy*, 2nd Ed. Santa Fe, NM: The Learning Zone.

Zeilik, M. 2003, Birth of the Astronomy Diagnostic Test: Prototest Evolution, *Astronomy Education Review*, 2(1), 46.

Zeilik, M., Bisard, W., & Lee, C. 2002, Research-Based Reformed Astronomy: Will It Travel?, *Astronomy Education Review*, 1(1), 33.

ÆR
178 - 186