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

Astronomy Education Review

Volume 2, Feb 2003 - Dec 2003
Issue 1

An Examination of Misconceptions in an Astronomy Course for Science, Mathematics, and Engineering Majors

by **Michael Zeilik**
University of New Mexico
Vicky J. Morris
University of New Mexico
Posted: 07/28/03

The Astronomy Education Review, Issue 1, Volume 2:101-119, 2003

Abstract

Using validated diagnostic questions (early versions of the Astronomy Diagnostic Test [ADT]), we identified misconceptions and tracked conceptual gains in two non-traditional college-level astronomy courses for science, mathematics, and engineering (SME) majors. We found large and robust gains overall, comparable to those in a non-traditional introductory course for non-SME students. We also discovered some disturbing misconceptions among the SME students, many of which are the same as those exhibited by non-SME majors. For both majors and non- majors, an attitude survey demonstrated a positive incoming belief that did not alter over one semester, but the SME majors did have higher (more positive) initial scores.

1. INTRODUCTION

Astronomy education research in higher education has focused on "Astro 101," the proverbial introductory survey course for non-science majors (Zeilik et al. 1997; Zeilik, Schau, & Mattern 1999). These courses have a diverse audience, as indicated by the self-reported majors of the students (Deming & Hufnagel 2001). To track conceptual growth in these courses—and the transformation of misconceptions to accepted ones—relies on field-tested, validated questions. In our early work, these items made up versions of the Astronomy Diagnostic Test (ADT 1.X), which contained a core of 19 questions (Zeilik, Schau, & Mattern 1998; Zeilik 2003). Some of this work was incorporated into ADT version 2.0, which has been used in a national survey in the United States (Hufnagel et al. 2000; Hufnagel 2002; Deming 2002). (Note: In this paper, "Astro 101" refers to the generic course; Astronomy 101 refers to a specific version of that course.)

In contrast, little work has been aimed at the introductory astronomy courses for science, mathematics, and engineering (SME) majors. The University of New Mexico (UNM) has a general astronomy course sequence for SME students that consists of two one-semester courses, Astronomy 270 in the fall and Astronomy 271 in the spring. Astronomy 270 focuses on the Sun and solar system; Astronomy 271 deals with the "rest of the universe." The level of mathematics employed in the course focuses on college algebra, trigonometry, and geometry; calculus is not used. Zeilik and Bisard (2002) gave initial results for these courses. Here, we present a comprehensive analysis for a two-year span (four semesters), including an assessment of attitudes in one semester. We believe that this is a first study of a neglected but vital audience. We were especially attentive to the misconceptions that such students bring to the course.

2. COURSE DESCRIPTIONS AND FORMAT

The UNM catalog describes these courses as dealing with "concepts of astronomy," with Astronomy 270 as the "recommended prerequisite" for Astronomy 271. In reality, about one- fourth of the students take the courses out of sequence, and only about half of the Astronomy 270 class continues into Astronomy 271 in the following semester—so, about half the class takes the courses out of sequence. Given this fact and that many concepts overlap, the instructor (MZ) decided to use an overtly conceptual approach to both courses rather than the mostly descriptive one used in the past (see Bisard & Zeilik 1998 for details). He modeled the conversion after that for the research-based version of Astronomy 101, in which we had developed and validated new assessment tools. (These tools are available on the Field-tested Learning Assessment guide [FLAG] at http://www.flaguide.org.)

These classes were small, typically about 30 students per semester. About 40% of the class reported that this was a course in their major field; another 40% were engineering majors for which the courses were suggested for fulfilling a science requirement. In addition, some 40% had an introductory college-level astronomy course. The highest levels of physics taken were: high school, 28%; college-level conceptual physics, 8%; college-level algebra-based physics, 16%; college-level calculus-based physics, 48%. The UNM catalog states that one semester of college-level algebra-based physics is a pre- or co-requisite for the courses, so about half of the students were "overprepared" in physics.

The results reported here are for two semesters of Astronomy 270 (fall 1996 and 1997) and Astronomy 271 (spring 1997 and 1998), which were taught using dedicated, heterogeneous, cooperative learning teams formed by the instructor for 50%-70% of class time. The instructor typically employed the rest for "mini-lectures" on specific topics—especially "hard" concepts—or linkages among concepts. These courses marked the first in a sequence of course innovations based on evidence from physics and astronomy education about "what works." Upon reflection, MZ considers these courses as "partially reformed" compared to the "Astro 101" courses that followed.

3. ASSESSMENTS AND RESULTS

The ADTs used to assess conceptual gains for this study consisted of 30 or 25 items in a pre/post experimental design. Fourteen of these items overlapped with those given in the ADTs for Astronomy 101. Because of the small numbers with voluntary participation, we have combined the results for both semesters: Table 1 for Astronomy 270 (ADT 1.5) and Table 2 for Astronomy 271 (ADT 1.6). (At UNM, the Human Subjects Committee requires that all assessments be voluntary and not connected to a course grade.) We checked the reliability of each test by calculating a standard internal reliability index called *Cronbach's alpha* (see Appendix A); ADT 1.5 had an index of 0.84 and ADT 1.6 had 0.83, both above the minimum level of acceptability (about 0.65). To quantify the conceptual change with respect to pre-test scores and possible gains, we calculated the **normalized gain index** (Hake 1998):

<*g*> = (%post - %pre)/(100 - %pre),

which can range from zero (no gain) to 1 (all possible gain). For Astronomy 270, <g(A270)> = 0.52 ± 0.16, and for Astronomy 271, <g(A271)> = 0.53 ± 0.19. The conceptual gains over one semester were the same for both classes.

Another figure of merit usually quoted in the realm of education research is the **effect size** (see Appendix A):

Effect size = ES = (post-test - pre-test)/ mean SD of the distributions

That is, the effect size is the difference between the means of the pre- and post-scores divided by their mean standard deviation, sometimes called "pooled standard deviations." Effect sizes of 1.0 or larger are considered to be outstanding (Cohen 1988). For Astronomy 270, ES (A270) = 2.44; for Astronomy 271, ES (A271) = 3.19.

Though very large, these effect sizes must be interpreted with caution because of the small number of participants involved. In addition, we calculated both the gains and effect sizes from the aggregate data, which tends to result in larger effect sizes.

We compared these results with those for Astronomy 101, a non-science majors course at UNM. In a voluntary sample of matched pairs (N = 586) for ADT versions 1.1, 1.2, 1.3, and 1.4 (given over four semesters), we found <g(A101)> = 0.49 and ES(A101) = 1.9. A better comparison shows up in the scores for the 14 items that Astronomy 101 and 270 tests had in common. On these, Astronomy 101 <g(A101)> = 0.53 ± 0.21, and Astronomy 270 <g(A270)> = 0.43 ± 0.24, statistically the same.

Table 1. Astronomy 270, Fall 1996 and 1997, UNM, N = 52, ADT 1.5

Test Item	**Pre-score (% correct)**	**Post-score (% correct)**	**Normalized gain index <g>**
1. Noon Sun	42	73	0.53
2. Moon phase/total solar eclipse	49	90	0.80
3. Angular size/distance	58	85	0.64

4. Sun's seasonal motion/horizon at sunset	60	72	0.30
5. Falling lead/wood spheres	86	95	0.64
6. Moon around zodiac	50	65	0.30
7. Radio vs. light waves	42	66	0.41
8. Inverse square law/light	40	69	0.48
9. Reasons for the seasons	39	62	0.38
10. Earth-Moon gravitation/Newton's 3rd law	25	49	0.32
11. Moon's monthly rotation/same side	33	52	0.28
12. Angular size/distance/Sun	76	78	0.08
13. Acceleration/mass	93	96	0.43
14. Acceleration/force	85	97	0.80
15. Model/Sun's motion through zodiac	18	64	0.56
16. Weight on Earth	61	75	0.36
17. Sun/close star scale model	61	74	0.33

18. Orion shape/distance	60	71	0.28
19. Jupiter’s moons/Kepler’s 3rd law	40	73	0.55
20. Inverse square law/gravitation	57	88	0.72
21. Luna Minor/centripetal acceleration	30	67	0.53
22. Luna Minor/ gravitational force	51	58	0.14
23. Moon/tidal forces	8	68	0.65
24. Escape speed/mass	52	73	0.44
25. Thermal emission/planet	62	91	0.76
26. Mass/planet’s evolution	36	72	0.56
27. Deimos/Phobos: Kepler’s 3rd/Mars’s mass	58	98	0.95
28. Conservation angular momentum: formation of solar system	38	88	0.81
29. Conservation of energy: formation of solar system	49	92	0.84

30. Sun's motion relative stars/horizon	28	82	0.75
Average ± SD	**50 ± 9.9**	**76 ± 8.4**	**0.52 ± 0.16**

Table 2. Astronomy 271, spring 1997 and 1998, UNM, N = 40, ADT 1.6

Test Item	**Pre-score (% correct)**	**Post-score (% correct)**	**Normalized gain index <g>**
1. Gravitational acceleration	90	100	1.0
2. Radio/visible light/same speed	44	68	0.43
3. Inverse-square/light	50	87	0.74
4. Newton's 3rd law	39	58	0.31
5. Angular diameter	72	90	0.64
6. Newton's 2nd law	88	100	1.0
7. Size & scale/Sun & nearby star	84	89	0.31
8. Stellar parallax/inverse proportion	45	58	0.24
9. Orion/viewpoint	60	84	0.6
10. Inverse-square law/gravitation	58	79	0.5

11. Centripetal acceleration	39	44	0.08
12. Gravitational force/masses	59	72	0.32
13. Tidal forces/inverse cube	20	38	0.22
14. Escape speed	62	84	0.58
15. Conservation angular momentum	74	92	0.69
16. Conversation of energy	52	90	0.79
17. Gravitational force/masses	61	76	0.38
18. Kepler's 3rd law	41	58	0.29
19. Stellar evolution	13	60	0.54
20. Angular diameter/distance	42	69	0.47
21. Flux/inverse square law/light	35	92	0.88
22. Tidal forces/black hole	26	44	0.24
23. Angular diameter/distance	45	70	0.64
24. Angular diameter/distance	76	92	0.67
25. Hubble law	78	94	0.73
Average ± SD	**54 ± 8.5**	**76 ± 4.2**	**0.53 ± 0.19**

4. MISCONCEPTIONS AND THEIR CHANGE

We now elaborate on items that revealed misconceptions, the most difficult concepts, and those that caused the most confusion in both classes. Tables 3 (Astronomy 270) and 4 (Astronomy 271) summarize these results for the pre-tests, which indicate the initial state of students' knowledge. Averaging these, we see robust trends emerge for these SME majors regarding physics and astronomy concepts.

Physics

1. 60% do not apply Newton's 3rd law to gravitational forces. Students consistently believe that a more massive object exerts a greater force on a less massive one.
2. 54% chose an inverse-square law for tidal forces rather than an inverse-cube relationship.
3. 42% indicate that escape speed depends on the mass of the escaping object, directly and as the square root of the mass.
4. 46% believe that different masses in the same orbit around a third mass have different centripetal accelerations, either directly proportional or inversely proportional to the orbiting masses (Newton's 2nd law).
5. 37% designate radio waves as traveling at a slower speed than visible light.
6. 30% select an inverse law for gravitation.
7. 23% select an inverse law for the flux of light.

(See Redish 2003 for a more complete review of common physics misconceptions.)

Astronomy

1. 47% believe that total solar eclipses occur at full moon.
2. 45% think that the Sun moves many degrees per day relative to the stars, confusing this with angular motion with respect to the horizon.
3. 42% deem that the Moon takes one year to circuit the zodiac, confusing the Moon's angular motion with respect to the stars to that of the Sun.
4. 39% judge that the Moon does not rotate on its axis from the observation that it keeps the same face toward the Earth.
5. 35% conclude that the Sun sets at the same position on the horizon a week after the fall equinox.
6. 30% deduce that seasons occur because the Earth's orbit is elliptical—hotter in summer because the Earth is closer to the Sun.

These lists overlap considerably with early ADT results from "Astro 101" at UNM and CMU, albeit with higher percentages of correct answers.

Table 3. Main misconceptions and confused concepts (20% or higher) from the ADT 1.5 pre-test using combined results; ranked from highest to lowest.

Test Item	Pre-score (%)	Misconception/confusion

10. Earth-Moon gravitation/Newton’s 3rd law	62	Gravitational force of Moon on Earth less than Earth on Moon
23. Moon/tidal forces	62	Inverse-square
21. Luna Minor/centripetal acceleration/same distance different mass as our moon	55	1/3 or 3
2. Moon phase/total solar eclipse	47	Full
24. Escape speed/mass	46	Depends on mass of escaping object
30. Sun’s daily motion relative to stars	45	Moves many degrees relative to stars
7. Radio vs. light waves	44	Radio travels slower
6. Moon around zodiac	42	One year
11. Moon’s monthly rotation/same side toward Earth	39	Does not rotate
4. Sun’s seasonal motion/horizon at sunset/fall equinox	35	Sunset at same place
22. Luna Minor/gravitational force	33	1/9 or 3
25. Thermal emission/planet	32	Convection
9. Reasons for the seasons	30	Earth closer in summer
20. Inverse square law/gravitation	29	Inverse
26. Mass/planet’s evolution	26	Distance from Sun

Table 4. Main misconceptions and confused concepts (20% or higher) from the ADT 1.6 pre-test using combined results; ranked from highest to lowest.

Test Item	**Pre-score (%)**	**Misconception**
4. Newton's 3rd law	55	Force of Earth on Moon greater than Moon on Earth
13. Tidal forces/inverse cube	46	Inverse square
22. Tidal forces/black hole	44	Inverse square
14. Escape speed	38	Depends on mass of escaping object
23. Angular diameter/distance/globular cluster	38	Unable to relate variables as inverse relationship
11. Centripetal acceleration	38	3 times or 1/3 as much
12. Gravitational force/different masses	36	Inverse square or no change
21. Flux/inverse square law/light	32	Inverse
2. Radio/visible light/same speed	30	Radio slower than visible
20. Angular diameter/distance/galaxies	30	Inverse square
10. Inverse-square law/gravitation	26	Inverse law
24. Angular diameter/distance/galaxies	25	Unable to relate variables as inverse relationship
3. Inverse-square/light	23	Inverse law

Now to look in detail at a few concepts selected because they had the most puzzling outcomes.

For the Astronomy 270 assessment (ADT 1.5), item #22 was:

An imaginary moon, Luna Minor, is one-third the mass of the moon and revolves around the Earth at the same distance as our Moon. What is the GRAVITATIONAL FORCE between the Earth and Luna Minor compared to the Earth and Moon?

A. Three times as much
B. Nine times as much
C. One-third as much
D. One-ninth as much
E. The same

Figure 1 shows each choice for this item, the fraction chosen by the participants, and the normalized gain index, <g>. Note the small change in <g> for choice *C*. Most of these came from choices *A* and *B*—these show negative gain indices. Choice *E* held steady. A slight gain occurred in *D*. MZ found these results strange and came to the following hypothesis: prior learning that emphasized the inverse-square law nature of Newton's gravitation interfered with the notion that the force is also directly proportional to the product of the masses. Clearly, instructional materials on this topic should be carefully designed so that students have the opportunity to understand both aspects of gravitation.

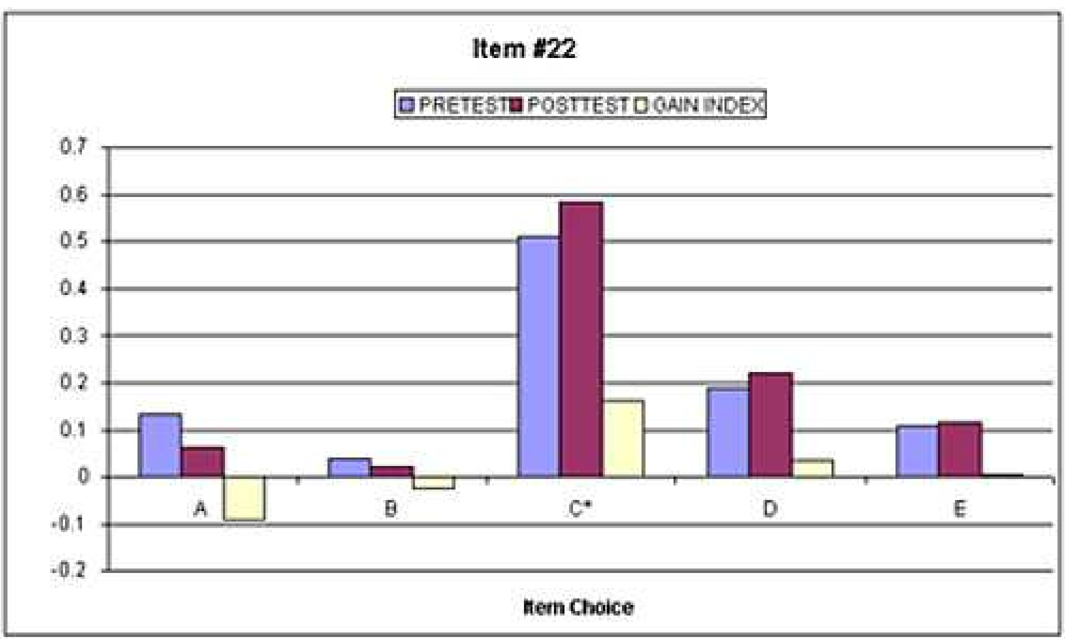

Figure 1. Pretest, post-test, and normalized gain index for Astronomy 270, Item #22 on ADT 1.5.

For the Astronomy 271 assessment (ADT 1.6), item #11 was:

An imaginary moon, Luna Minor, is one-third the mass of the Moon and revolves around the Earth at the same distance as our Moon. What is the CENTRIPETAL ACCELERATION between the Earth and Luna Minor compared to the Earth and Moon?

A. Three times as much
B. Nine times as much
C. One-third as much
D. One-ninth as much
E. The same

Figure 2 shows each choice for this item, the fraction chosen by the participants, and the normalized gain index, <g>. Note the small change in <g> for choice *E*. Most of these came from students selecting choices *C* and *D*—these show negative gain indices. Choice*A* held steady. A slight gain occurred in *B*. We believe that these results show that, although a substantial fraction of students entered with a basic understanding of the concept, prior learning impeded large gains by other students.

MZ was puzzled by these results because he felt that much instructional time and effort was devoted to these topics. On the other hand, confusion between force and acceleration is common in all levels of introductory physics courses (Redish 2003).

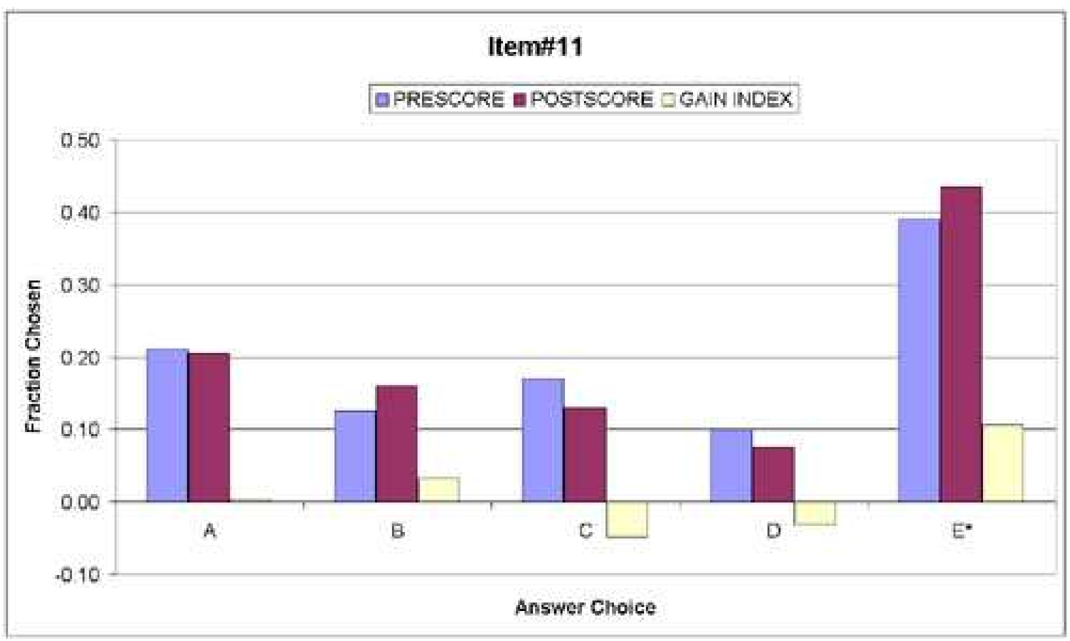

Figure 2. Pretest, post-test, and normalized gain index for Astronomy 271, Item #11, ADT 1.6.

Also in ADT 1.6 for Astronomy 271 was an item about tidal forces:

Imagine that our Moon is moved to three times its current distance from the Earth. How much would the Moon's TIDAL FORCES then compare to those on the Earth now?

A. The same
B. 1/3
C. 1/9
D. 1/27

This item probes student understanding of tidal forces as differential gravitational forces that change as the inverse cube of the distance. (Note that this concept does not require calculus to derive the inverse-cube result.) Figure 3 shows the results. Answer *D* achieved a gain of only 0.22, most of which came out of answer *B* (note the negative gain). Answer *C* showed a slight gain. Most disheartening to the instructor was the result that about 46% of the ADT participants thought the tidal forces were proportional to the inverse square of the distance, while only 38% choose an inverse-cube relationship. Again, the instructor subjectively believed that this topic was allocated considerable instructional time and emphasis and believes that an "interference effect" from prior knowledge and instruction took place again. Historically, these concepts are difficult, even for the self-selected population enrolled in these courses.

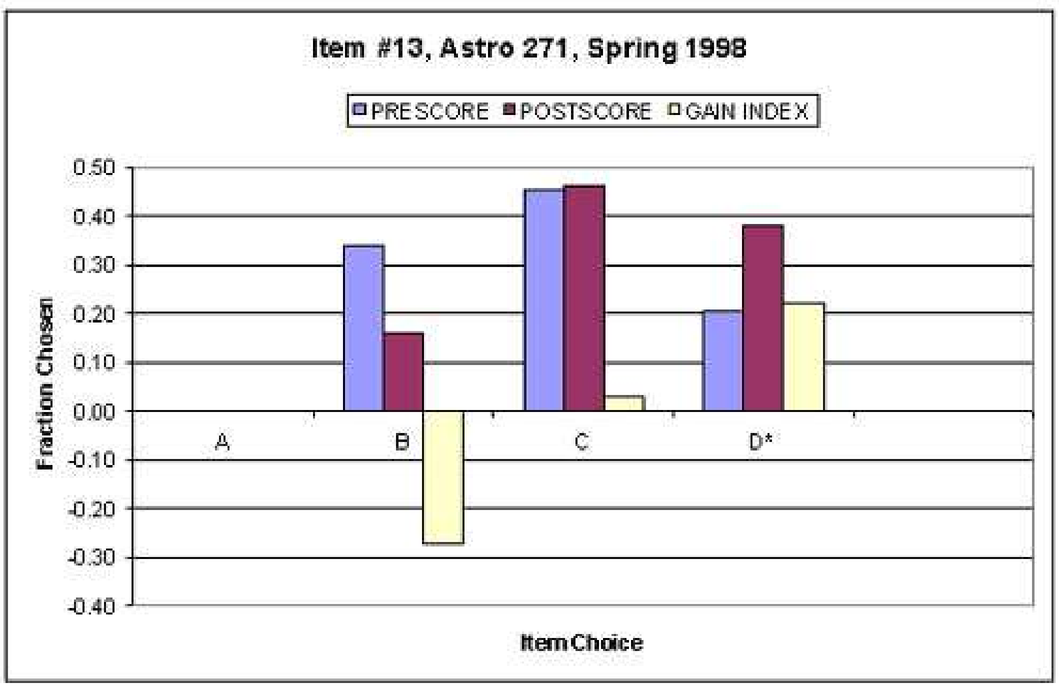

Figure 3. Pretest, post-test, and normailized gain index for Astronomy 271, Item #13, ADT 1.6.

5. ATTITUDE SURVEY

For Astronomy 271, we gave the students an attitude survey (Attitude Survey II in Zeilik, Schau, & Mattern 1999). The survey consists of 34 items divided into four subscales:

1. **Affect**: positive and negative feelings concerning astronomy/science
2. **Cognitive competence**: attitudes about intellectual knowledge and skills when applied to astronomy/science
3. **Value**: attitudes about the usefulness, relevance, and worth of astronomy/science in personal and professional life
4. **Difficulty**: attitudes about the difficulty of astronomy/science as a subject

Each item was on a five-point Likert scale so that 3 = the middle value (neither disagree or agree), 1 = strongly disagree, and 5 = strongly agree. (This survey is available at www.flaguide.org.)

Overall, pre-test results (N = 28) were:

Mean (pre) = 3.92 ± 0.55 (SD)

and the post-test results (N = 15) were:

Mean (post) = 4.14 ± 0.61 (SD)

Comparing the pre- and post-test results using a standard statistical test (a *t-test*; see Appendix A), we found no statistically significant difference between the scores, although the trend is upward. We also discovered the same to be true for the subscales (see Table 5). Not surprisingly, the class started with a strong positive attitude toward science in general and astronomy in particular. The pre-test result was no revelation given the content of the course and the academic backgrounds and interests of the students. What surprised us was that the mean class score did not change over one semester.

This "no change" outcome mirrors that for UNM’s Astronomy 101 in fall 1995, where we found a stable, slightly positive attitude (mean = 3.14 ± 0.37 for N = 108 matched pairs) that did not vary over one semester (Zeilik, Schau, & Mattern 1999, Table IV). The same held for "Astro 101" students at CMU when measured by the same attitude survey (Zeilik, Bisard, & Lee 2002): mean (pre-test, N = 237) = 3.17 ± 0.30; mean (post, N = 224) = 3.10 ± 0.32.

These results should give pause to introductory astronomy instructors for whom enhancing attitudes is a major affective outcome. That certainly is a worthy goal and one that MZ supported for many years. But the reality is that over one semester, our students’ attitudes as a whole do not seem to change as measured by our assessment tool. To the best of our knowledge, we have conducted the only assessments of attitude change in Astronomy 101 and SME majors at the same institution. The Astronomy 101 results start somewhat positive and do not change significantly. The same is true for the SME majors course at UNM, using the same validated instrument. Note that the levels of the SME results are larger than those for Astronomy 101, as would be expected in the different audiences.

Two particular items did exhibit statistically significant changes that were masked in the mean scores for the SME majors. "Astronomy concepts are easy to understand" (cognitive competence subscale) showed a difference at the $p \leq 0.021$ probability level (interpreted as 2.1% that the difference is chance), and "Scientific skills will make me more employable" (value subscale) differed pre/post at the $p \leq 0.042$ level (4.2% that the difference is chance). The usual standard for this type of test is $p \leq 0.05$. We interpret this test cautiously because we prefer results at $p \leq 0.01$ or better.

Table 5. Attitude survey results, Astronomy 271, spring 1998

Subscale	Mean ± SD (pre)	Mean ± SD (post)	Statistical significance*
Value	4.26 ± 0.82	4.43 ± 0.53	NS
Difficulty	3.00 ± 0.55	3.36 ± 0.68	NS
Cognitive competence	4.08 ± 0.66	4.36 ± 0.78	NS
Effect	4.27 ± 0.69	4.33 ± 0.85	NS
All	**3.92 ± 0.55**	**4.14 ± 0.61**	**NS**

*Based on an independent samples t-test (Appendix A)

6. IMPLICATIONS

We have found no published report of a conceptual assessment for a two-semester introductory astronomy course aimed at an SME audience. For contrast, then, we will present the results with similar, validated instruments from Astronomy 101 classes at UNM. These courses clearly sample different populations that are worthy of comparison if we desire to highlight conceptual differences between them.

We found that both populations struggle with similar concepts that they reported as "difficult." The Astronomy 101 group starts at a lower conceptual understanding as measured by the ADT 1 (pre-test means typically 35%-45%). These students and those in the SME course, however, exhibit the same size conceptual gains as measured by a normalized gain index ($\langle g \rangle \approx 0.5$) or an effect size (ES ≈ 2).

This experiment did not separate the effect of cooperative learning alone from that of the overall conceptual design. A recent review of cooperative learning in SME courses concludes conservatively that the effect size on achievement is 0.5 (Springer, Stanne, & Donovan 1999). We recurrently get effect sizes of ≈ 2 for reformed astronomy courses. We consider the Springer et al. (1999) results to fix a lower bound. We attribute these gains (above that for cooperative learning alone) to the conceptual emphasis of the courses and assessments and achievement quizzes/tests that align with this emphasis.

What can we infer from these results? First, that incoming SME students (at least at UNM) enter a course with a set of misconceptions similar to that of "Astro 101" students. They also exhibit major misconceptions on advanced concepts appropriate to the higher course level. *Instructors must measure this prior knowledge to develop effective teaching strategies*.

Second, we deduce that even in a partially reformed astronomy course with a strong conceptual emphasis and cooperative learning*, conceptual gains assessed by a validated conceptual diagnostic test are comparable in large or small one-semester classes with diverse student populations*. We can improve learning in such courses if we apply active learning formats shaped by the results of disciplinary-based education research.

Acknowledgments

This work was supported in part by NSF grant DUE-9981155. We thank Tim Slater and M. Jennifer Markus for a close reading of the manuscript.

Copies of the assessments in PDF can be requested from MZ by e-mail: zeilik@la.unm.edu.

Appendix A

Brief on Educational Statistics Used in this Paper

Cronbach's alpha

Cronbach's alpha measures how well a set of items or variables measures a single one-dimensional latent construct. When data have a multi-dimensional structure, Cronbach's alpha usually will be low. Technically speaking, Cronbach's alpha is not a statistical test; it is a coefficient of reliability or consistency.

Cronbach's alpha can be written as a function of the number of test items and the average inter-correlation among the items. For conceptual purposes, the formula for the standardized Cronbach's alpha is:

Alpha = N x r/[1 + (N-1) x r]

Here, *N* is equal to the number of items and *r* is the average inter-item correlation among the items.

You can see from this formula that if you increase the number of items, you increase Cronbach's alpha. In addition, if the average inter-item correlation is low, alpha will be low. As the average inter-item correlation increases, Cronbach's alpha increases as well.

This makes sense intuitively. If the inter-item correlations are high, then there is evidence that the items are measuring the same underlying construct. This is really what is meant when someone says they have "high" or "good" reliability. They are referring to how well items measure a single one-dimensional latent construct.

t-test

If we are comparing only two means of test scores, we can use a *t-test*. The t-test is the most frequently used inferential statistics test to check the statistical probability that the means from two samples come from populations with identical means. A statistically significant *t-value* indicates that a mean difference of this size would have occurred due to sampling error (chance) at the probability level (*p*-value) associated with the specific t-test value. When that probability is small (equal to or less than 0.05, or 5%), we can conclude that the means likely differ. Using 5%—a typical choice—we have a 95% chance of being correct in this judgment.

Effect size

Effect size is the difference between two means in standard deviation units. In essence, it measures the average superiority (if positive) or inferiority (if negative) of the final state compared to the initial state, while taking into account the variability of the population. Effect size is a powerful indicator of the separation of the pre- and post-course score distributions, and so of the gains across a span of time. It permits a calibration of comparisons across different characteristics of a study by normalizing the results by standard deviations. In the educational research literature, effect sizes of 0.1 or less are considered small and of no practical import; 0.3, medium and of practical significance; and 0.5 or greater, large and unusual effects (Cohen 1988).

References

Bisard, W., & Zeilik, M. 1998, Restructuring a Class, Transforming the Professor: Conceptually Centered Astronomy with Actively Engaged Students, *Mercury*, 27 (4): 16.

Cohen, J. 1988, *Statistical Power Analysis for the Behavioral Sciences*, Mahwah, NJ: Lawrence Erlbaum.

Deming, G. 2002, Results from the Astronomy Diagnostic Test National Project, *Astronomy Education Review*, 1 (1), 52. http://aer.noao.edu/AERArticle.php?issue=1&section=1&article=5.

Deming, G., & Hufnagel, B. 2001, Who's taking ASTRO 101?, *The Physics Teacher*, 39, 368.

Hake, R. R. 1998, Interactive-Engagement vs Traditional Methods: A Six-Thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses, *American Journal of Physics*, 66 (1): 64. http://www.physics.indiana.edu/~sdi/.

Hufnagel, B. 2002, Development of the Astronomy Diagnostic Test, *Astronomy Education Review*, 1 (1), 47. http://aer.noao.edu/AERArticle.php?issue=1&section=1&article=4.

Hufnagel, B., Slater, T., Deming, G., Adams, J., Lindell, R., Brick, C., & Zeilik, M. 2000, Pre-Course Results from the Astronomy Diagnostic Test, *Pub. Astron. Soc. Australia*, 17:2. http://www.atnf.csiro.au/pasa/17_2.

Redish, E. F. 2003, *Teaching Physics with the Physics Suite*, New York: John Wiley & Sons.

Springer, L., Stanne, M. E., & Donovan, S. S. 1999, Effects of Small-Group Learning on Undergraduates in Science, Mathematics, Engineering, and Technology, *Review of Educational Research*, 69 (1), 21.

Zeilik, M. 2003, The Birth of the Astronomy Diagnostic Test: Prototest Evolution, *Astronomy Education Review*, 1 (2), 46. http://aer.noao.edu/AERArticle.php?issue=2&section=2&article=5.

Zeilik, M., Schau, C., & Mattern, N. 1998, Misconceptions and Their Change in University-Level Astronomy Courses, *The Physics Teacher*, 36, 12.

Zeilik, M., Schau, C., Mattern, N., Hall, S., Teague, K. W., & Bisard, W. 1997, Conceptual Astronomy: A Novel Model for Teaching Postsecondary Science Classes, *American Journal of Physics*, 65, 987.