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

Astronomy Education Review

Volume 1, Oct 2001 - Jan 2002
Issue 1

Research-Based Reformed Astronomy: Will It Travel?

by **Michael Zeilik**
University of New Mexico
Walter Bisard
Central Michigan University
Carl Lee
Central Michigan University
Posted: 05/31/02

The Astronomy Education Review, Issue 1, Volume 1:33-46, 2002

Abstract

This study investigated the transfer of a research-based conceptual course in university-level introductory astronomy from its original setting to another institution. The primary implementation took place at the University of New Mexico (UNM), the secondary one at Central Michigan University (CMU), both minimally funded state institutions with scarce teaching resources. The experiment intended to replicate as many aspects of the primary implementation as possible, while controlling for the use of cooperative learning teams.

Overall, the CMU results over one semester showed large conceptual gains--close to those found at UNM on a concept map assessment and somewhat lower than UNM on an astronomy diagnostic test. No differences appeared between the two CMU sections, one that used cooperative learning teams and one that did not. Within both CMU sections, males outscored the females on an astronomy diagnostic test, but the gains of both groups were the same. Attitudes, which were somewhat positive on the pre-assessment, were unchanged at both UNM and CMU on the post-assessment. We conclude that that the UNM reformed model can be transferred to a different institutional context and instructor with reasonable success. Instructors using this model can expect robust conceptual gains, even if they do not implement cooperative learning teams.

1. INTRODUCTION

Many colleges and universities in the U.S. offer a large-enrollment, introductory astronomy class (Fraknoi 1996, 2001). Such courses, aimed at nonscience majors, attempt to "cover" the Universe in one semester. They are normally taught as lectures, with generous audiovisual and some computer-assisted modes of instruction (Fraknoi 1996, 2001). Instructors overwhelmingly practice direct didactic lecture with little attempt to engage students actively. What impact do these survey courses have on students?

Emerging astronomy education research (Pasachoff & Percy 1990; Gouguenheim, McNally, & Percy 1998; Sadler 1987; Bisard et al. 1994; and others) demonstrate the lack of effectiveness in learning physics and astronomy concepts compared to instructors' expectations in introductory courses. Pertinent results from physics education research show that conceptual learning in a traditional lecture format is *not* affected by the skills and style of the lecturer, use of audio-visuals, or the implementation of computer technology (for example, Van Heuvelen 1991; McDermott 2001; Sadler 1992; Redish 1994; Halloun & Hestenes 1985; McDermott 1984). Interactive methods in physics instruction have generally shown larger gains on standardized physics mechanics tests (as summarized by Hake [1998]; see Redish & Steinberg [1999] for a synthesis and McDermott & Redish [1999] for an overview of research on learning physics conceptually).

The lack of significant conceptual change in large, traditional science classes may well relate to misperceptions of the students that, in the case of astronomy education, result in a "triple whammy" as pointed out by Zeilik & Bisard (1997), Zeilik et al. (1996), Zeilik et al. (1995), Zeilik (l996):

- Students view astronomy as a narrow, observationally based subject dealing chiefly with constellations, planets, NASA, black holes, and cosmology;
- Students have prior science misconceptions that impede learning astronomy concepts;
- Students view science as a large assembly of unrelated facts to be memorized.

These misperceptions create in students the expectation that science courses will deliver concepts in an unconnected way ("memorizing a massive collection of facts"). Yet a true conceptual understanding of astronomy concepts must involve rich connections, similar to those of experts. Our reformed course model aims to advance these connections, the "big picture" of astronomy.

In our previous work (Zeilik et al. 1997), we presented the cognitive basis of a reformed course model, its implementation at UNM, and an analysis of the results. We found large gains in conceptual understanding in the first semester of full implementation (Fall 1995), which built upon four semesters of development. Follow-up research (Zeilik, Schau, & Mattern 1999) confirmed our positive results over three semesters with hundreds of students. This project pioneered a large-scale, long-term study of an introductory course based on "best practices" within a cognitive model applied to a large-class setting.

But will the model work in a different context? Here we examine the transfer of the course to another institutional setting with an instructor other than the originator. Such transfer issues are crucial to the spread and success of research-based learning innovations in higher education. Without proven portability, reform will fail. Rarely have studies been carried out in astronomy or physics education research that address the question, Will it travel?

2. MODEL IMPLEMENTATION AT CENTRAL MICHIGAN UNIVERSITY

Our reformed model incorporates a conceptually based class integrated with instructional and assessment tools to evaluate the course in a resource-poor environment. Our approach differs from the usual application of physics education research. In these, typically one aspect of the course is changed (such as discussion sections or labs), while the main lecture is left untouched. Both UNM and CMU lack discussion sections and a required laboratory for all students enrolled in the lecture sections. Hence, all innovations must be applied in the lecture class, with minimal aid (if any) from a teaching assistant. The reality of our institutions forced us to take an integrated view and to apply best practices to create active learning (Bonwell & Eison 1991; Meyers & Jones 1993; Mazur 1997; Mintzes, Wandersee, & Novack 1998) environments in a resource-poor context.

The key feature of the transformed course resides in its intense focus on connected understanding of astronomy. Hence, the central instructional strategy employed concept mapping (Novak & Gowin 1984; Schau, Mattern, & Teague 1999) as a visual guide to the connected understanding of the concepts in the lessons, readings, and activities. A concept map is a diagram--often hierarchical--of nodes, each containing concept labels, which are linked by directional lines, also labeled. Instructional concept maps depicting about 100 essential concepts play a key role in designing the conceptual astronomy format at UNM (Zeilik et al. 1997). There we developed three global pre- and post-Select and Fill-In (SAFI) concept maps to assess the acquisition of connected knowledge. The study at UNM demonstrated that the SAFI instrument has a post-test reliability of about 0.8 (Zeilik, Schau, & Mattern 1999). We measured reliability by a standard statistical index that is the average intercorrelations of all items in the measure. A value of 0.65 is generally considered the lower limit of acceptable reliability.

The conventional achievement measurements were the conceptually based multiple choice tests and quizzes used in both sections. These tests were designed to probe applications or extensions of concepts rather than recall ability. However, it is difficult to test students' understanding of connected understanding with such items. For this reason, we used the SAFI concept maps to assess students' broad understanding on how the concepts are connected.

A measure of students' understanding on particularly resilient misconceptions in astronomy was also employed to assess the gains of learning over the course. Our misconceptions measure had a multiple-choice format, with concepts and distracters based on the misconceptions research in physics and astronomy (Zeilik, Schau, & Mattern 1988; Zeilik & Bisard 2000). Each item has been statistically monitored to yield reliable measures over several semesters (post-test reliability of about 0.7). These items comprised the Astronomy Diagnostic Test, version 1.3 (ADT 1.3).

This report presents a culmination in one semester of the complete implementation at CMU with all pre/post assessment instruments. During the spring semester of 1996 at CMU, the morning class (11:00 a.m. to noon) of 101 students met three days a week and adopted a direct instruction approach (modified by the conceptual underpinnings). The afternoon class (2:00 to 3:20 p.m.) of 181 students met two days a week on Mondays and Wednesdays and involved dedicated cooperative learning teams. These cooperative teams were randomly selected alphabetically from the class list, with four students in some 50 teams. The groups followed research-based cooperative learning techniques (Johnson, Johnson, & Smith 1991; Michaelsen 1992; Michaelsen, Black, & Fink 1996; Millis & Cotell, 1998), with at least one activity per week. The same instructor (WB) taught both class sections. The experiment difference was this use of cooperative learning teams in the afternoon class. The group activities were targeted at particularly difficult concepts that were identified by research and the experience of the instructor. They demanded group discussion, written consensus, and completion in roughly 30 minutes of class time.

The two sections did *not* differ significantly in any demographic factors or from the previous three semesters of research with CMU students in the same course. Also, no evidence indicates differences between sections because of the time of day or number of sessions per week. The two sections were well matched.

The self-reported ethnicity was largely Anglo-American (95%) with very few minorities, equally divided with respect to gender. The largest major group was education (38%), followed by business (25%) and science/math (20%). Nearly 35% had taken a high school physics course. About 60% had studied earth science in high school, and a similar fraction had taken at least high school geometry. Most students used computers for word processing, and a simple majority felt "average" in science. Only 20% were currently taking the astronomy laboratory course. About 45% of the students expected to spend 3 to 4 hours a week studying for the course, while another 30% expected to study 5 to 6 per week.

3. RESULTS

Large-enrollment classes at a regional, resource-poor university such as CMU dictated that full implementation occur over three semesters of gradual execution and testing. Here, we first compare results between the CMU sections and then between CMU and UNM.

Four achievement instruments were used to investigate differences between the CMU sections. The assessment data were: (1) average score on four tests; (2) final class score, which determined the students' grade; (3) posttests of the three concept maps; and (4) ADT 1.3 posttests. Generally, students' achievements were *indistinguishable* between the two CMU classes (based on ANOVA tests; see Appendix A).

Two conceptual instruments had a pre- and posttest format: the global concept maps and the ADT 1.3. We measured the degree of gain by comparing pretest average/posttest average for each class. Table 1 summarizes the results for ADT 1.3. We compared pre- and post-course results using a normalized gain index, <g>, the ratio of the actual average gain to the maximum possible average gain:

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

A <g> of 0 means no gain, while a <g> of 1 indicates that all possible gain occurred (Note 1).

The mean CMU gains for the ADT 1.3 was <g> = 0.27; the difference between the two sections was *not* statistically significant. The mean gains on the concept maps also showed *no* statistically significant difference, with <g> = 0.32 for both sections combined (see Table 2).

Table 1. Results from ADT 1.3 for introductory astronomy courses at CMU (Spring 96) and UNM (four semesters, from Fall 94 to Spring 96). Based on a dependent-means t- test, all pre- post differences were statistically significant at the $p < 0.001$ level or smaller for the aggregated data (Appendix A).

Measure: ADT 1.3	Pretest mean (% correct ± SD)	Posttest mean (% correct ± SD)	Gain (*g*) and effect size (ES)
CMU coop	45.6 ± 13.2 (N = 185)	59.2 ± 14.4 (N= 179)	*g* = 0.25 ES = 1.10
CMU non-coop	43.6 ± 12.4 (N = 109)	60.4 ± 26.6 (N = 103)	*g* = 0.30 ES = 1.20
All CMU	44.9 ± 12.9 (N = 294)	59.6 ± 19.2 (N = 282)	*g* = 0.27 ES = 1.15
All UNM coop	38.1 ± 13.0 (N = 602)	67.5 ± 17.2 (N = 594)	*g* = 0.48 ES = 2.09

Table 2. Results from Select and Fill-In Concept Maps for introductory astronomy courses at CMU (Spring 96) and UNM (four semesters, from Fall 94 to Spring 96). Based on a dependent means t-test, all pre-post differences were statistically significant at the $p < 0.001$ level or smaller for the aggregated data (Appendix A).

Measure: Concept Maps	Pretest mean (% correct ± SD)	Posttest mean (% correct ± SD)	Gain (*g*) and effect size (ES)
CMU coop	36.2 ± 15.5 (N = 116)	53.4 ± 23.1 (N= 125)	*g* = 0.27 ES = 0.89
CMU non-coop	34.8 ± 15.5 (N = 84)	55.5 ± 22.4 (N = 82)	*g* = 0.32 ES = 1.09
All CMU	35.6 ± 15.5 (N = 200)	56.1 ± 22.8 (N = 207)	*g* = 0.28 ES = 0.97
All UNM coop	30.0 ± 11 (N = 441)	50.3 ± 19 (N = 594)	*g* = 0.29 ES = 1.22

We can also track the gains using an *effect* size parameter. Conceptually, the effect size of a statistically significant result is the difference between the means of the post- and pre-course score distributions, normalized by their mean standard deviation (Appendix A). The effect size at CMU (Table 1) for ADT 1.3 averaged 1.15, and for concept maps, 0.82, with *no* significant differences between the two sections. If we assume a normal distribution for the pretest and posttest scores, then an effect size of 1.15 means that almost 90% of the post scores exceeded the mean of the pretest scores; one of 0.82 that 80% did so. These are very large gains in conceptual knowledge by educational research standards.

These results suggest that students with various backgrounds in each class at CMU had similar gains by the end of the semester, based on two non-traditional instruments, and that each section had similar gains. The exceptions were: (1) High school (especially calculus) and college math backgrounds, with more math having a weak positive correlation; (2) Gender, with males outscoring females.

Point 2 is a major equity issue. Females had an ADT 1.3 posttest mean of 49% and a mean test score of 67%; in contrast, males scored 55% and 70%. Though women on average scored lower than men on the misconception posttest, their pre/post gain was essentially the same as that of the men (<g> = 0.55 vs. 0.50).

We anticipated that a student with higher scores on the achievement test should also have a "more connected" knowledge of concepts as examined by the concept maps and ADT 1.3. The relationship between each pair of achievement instruments provides insight of students' achievement from various dimensions. Table 3 summarizes the correlation between each pair of achievement scores for final scores, test average scores, and the post scores on the concept map and ADT 1.3 for both CMU sections combined.

Table 3-A. Correlation Between Measures: CMU Non-Coop Section

	Four Test Average	**Concept Map Posttest**	**Misconception Posttest**	**Attitude Posttest**
Final Score	.887* .000†	.496 .000	.594 .000	.408 .000
Four Test Average		.511 .000	.662 .000	.493 .000
Concept Map Posttest			.332 .003	Not significant
ADT 1.3 Posttest			1.0 .000	.361 .000

* The correlation coefficient, *r*.
† The *p*-value of the t-test (Appendix A).

Table 3-B. Correlation Between Measures: CMU Coop Section

	Four Test Average	Concept Map Posttest	Misconception Posttest	Attitude Posttest
Final Score	.728* .000†	.474 .000	.423 .000	.247 .006
Four Test Average		.554 .000	.620 .000	.394 .000
Concept Map Posttest			.316 .000	.187 .042
ADT 1.3 Posttest			1.00 .000	.359 .000

* The correlation coefficient, *r*.
† The*p*-value of the t-test (Appendix A).

As expected, the two highest related instruments were final score and test average in both classes, because the final score consisted chiefly of the four test scores. For other pairs of instruments, every pair was significantly correlated. As a general guideline in educational research, a correlation coefficient of 0.1 or less is considered small in practice; around 0.3 medium; and around 0.5 or more, large (Cohen 1988). The CMU results were medium to large. More impotant, the variances of concept map and ADT 1.3 scores with the four- test average ranged from 0.18 to 0.35, which indicates that these instruments probed overlapping domains of knowledge.

How did the CMU results compare to those at UNM? The two courses were mismatched demographically (Note 2). The largest differences between CMU and UNM students were ethnicity, age, and major. UNM students were more diverse ethnically, somewhat older in age, and included far fewer education majors. About 30% were Hispanic American, and some 80% expected to work on the course up to 5 hours per week. About 70% were under 22 years of age (which is atypical of UNM, where the mean undergraduate age was 27). All the UNM sections used cooperative learning, and they all met twice a week in 75-minute sessions. Very few UNM students (a few per class, about 1%) were education majors.

To augment the conceptual and traditional assessments, the UNM project developed two attitude surveys. The second of these--called Attitude II--was given to the CMU sections. Attitude Survey II contained 34 items and originally used a 1 to 7 Likert scale (strongly agree to strongly disagree), so that 4 was a neutral score. A modified version was used at CMU; it had a 1 to 5 scale, so that 3 was neutral. Both measures were scored such that higher scores always indicated more favorable attitudes, even though some of the items were negatively worded--a common practice for survey items to minimize bias.

To make the total attitude scores comparable, we subtracted the response to each item from the scale score representing the most negative attitude (1 for both surveys) and then divided that difference score by the total possible spread (4 points on 5-point scale and 6 on 7-point scale) to give the relative position on the scale. The resulting score is the proportional location on the attitude scale, with 50% as neutral (Table 4).

The CMU results show *no* attitude changes are measured across a semester for both classes. Note, though, that both showed somewhat positive attitudes entering the course, and these held up over one semester. The same was true for attitudes measured at UNM.

Table 4. Results from Attitude Survey (version II) for introductory astronomy courses at CMU (Spring 96) and UNM (two semesters, from Fall 95 to Spring 96). Scores have been normalized so that 50% is a neutral response. Based on a dependent means t-test, all pre/post differences were statistically significant at the $p < 0.01$ level or smaller for the aggregated data (Appendix A).

Measure: Attitudes II	**Pretest mean score (%)**	**Posttest mean score (%)**	**Gain (*g*) and effect size (ES)**
CMU coop	56 (N = 146)	55 (N= 138)	Not statistically significant
CMU non-coop	58 (N = 91)	56 (N = 86)	Not statistically significant
All CMU	57 (N = 237)	55 (N = 224)	Not statistically significant
All UNM coop	60 (N = 207)	58 (N = 184)	Not statistically significant

On the concept map measure (Table 2), the gains at the two institutions were essentially the same ($\langle g \rangle \approx 0.3$). This result reinforces our conviction that even novices can attain connected understanding in a field if it is taught explicitly.

Finally, the gains on the ADT 1.3 differed, with UNM showing significantly higher gains than CMU ($\langle g \rangle = 0.48$ vs. $\langle g \rangle = 0.27$). We do not have a clear explanation for this difference from this experiment.

4. SUMMARY AND IMPLICATIONS

The application of a conceptual paradigm to a large introductory astronomy course at CMU resulted in supporting most pedagogical strategies. We emphasize that, by design, *both* sections had a strong conceptual foundation, especially in the use of instructional concept maps to make connections between concepts explicit--a development that took place over three semesters with hundreds of students before the experiment was conducted. However, this was *not* a physics experiment, even though we used "best efforts" to "control" variables in the CMU sections (such as instructor, textbook, course content, tests, and grading). We intended cooperative learning--as the treatment in one section and not in the other--to serve as the independent variable within the CMU context. All had to be done in a naturalistic setting, which limits variable "control."

In retrospect, we believe that the overall conceptual organization and processes used in *both* sections may have largely washed out the effects of cooperative learning by itself. Generally, the successful implementation of cooperative learning *by itself* results in gains with effect sizes of at least 0.5 (Note 3). The process of stressing the conceptual connections with concept maps and other means before, during, and after all instruction marked the major difference of this version of the CMU astronomy course compared to a direct instruction, descriptive format.

Reflecting upon our results and experiences, we offer these main observations:

- Both CMU sections showed large conceptual gains, as gauged by the ADT 1.3 and concept map assessments; but no differences emerged between the sections on these measures.
- The most consistent CMU demographic effect in this study was gender, with males scoring higher on the four classroom tests (70% vs. 67%) and the ADT 1.3 (55% vs. 49%); but women had similar gains as the men on the concept maps and the ADT 1.3.
- The effects of prior schooling in the physical sciences were undetectable and in mathematics were small; a weak correlation was found between achievement and high school or college calculus.
- Both sections showed somewhat positive attitudes toward science and astronomy upon entering the course; these attitudes did not change over one semester.
- Comparing CMU and UNM results, we note that concept map gains and attitude changes were the same; UNM outgained CMU on the ADT 1.3.

From the personal view of the professor (WB), the afternoon CMU astronomy class with interactive techniques was much more fun and lively to teach: The students came to class more often, arrived early, and asked many more questions than the students in the more traditional section. Moreover, the professor did cover the same amount of material in both sections, and didactic presentations were much more "connected" because of the instructional design (Bisard & Zeilik 1998).

Rare are published reports of the transfer of innovations (Note 4). We know of none in astronomy, but a few in physics. McDermott & Reddish (1999) do refer to Physics by Inquiry and Workshop Physics, but a specific section on dissemination is lacking. Redish & Steinberg (1999) point out the large normalized gain indices (and narrow spread, compared to traditional methods) for these two research-based curricula. Saul and Redish (1998; also see Redish & Steinberg 1999 and Redish 1999) examine the dissemination of Workshop Physics for 11 institutions with secondary implementations. They conclude, based on Force Concept Inventory (FCI) normalized gains, that the secondary implementations were comparable to those at the primary site (Dickinson College). Cummings et al. (1999) evaluated Studio Physics at Rensselaer Polytechnic Institute and compare it to Interactive Lecture Demonstrations and Cooperative Group Problem Solving at other institutions. They found the Studio Physics gains, measured by the FCI and the Force and Motion Conceptual Evaluation, were half that of the other strategies and comparable to that for a traditional physics class. Our outcomes matched those for Workshop Physics' dissemination: the results of secondary implementation (CMU) were comparable to those of the primary implementation (UNM).

Pooling the CMU and UNM data, we have eight semesters of experience and over one thousand students in reformed astronomy classes. We find large and robust conceptual gains at both institutions. We conclude that the integration of "best practices" in an active learning format, coupled with alternative assessments, creates an effective and robust learning environment in large classes. We accomplished this goal at two minimally funded state institutions with very limited instructional resources. The transfer of institutional setting was largely successful. Instructors who wish to adopt and adapt our model can test for its portability in the context of their own courses and institutions using our assessment tools (Note 5). They can expect large conceptual gains by their students, even if they do not use cooperative learning teams.

Acknowledgment

This research was supported in part by NSF grants No. DUE-9253983 and DUE-9981155 to MZ.

Notes

1. Hake defined the normalized gain in this way; see Hake, R. R. 1998, "Interactive-engagement vs Traditional Methods: A Six-thousand-student Survey of Mechanics Test Data for Introductory Physics Courses." *Am. J. Phys.* 66(1):64-74; online at http://www.physics.indiana.edu/~sdi/. He later discovered an earlier use in Hovland, C. I., Lumsdaine, A. A. & Sheffield, F. D. 1949, "A Baseline for Measurement of Percentage Change," in *Experiments on mass communication*. C. I. Hovland, A. A. Lumsdaine, and F. D. Sheffield (Editors), Wiley 1965 (first published in 1949). Reprinted as pages 77-82 in "The Language of Social Research: A Reader in the Methodology of Social Research," P. F. Lazarsfeld and M. Rosenberg (Editors), 1955, Free Press, New York.

2. The UNM demographics are *very* different from those in a national survey with ADT 2. In that sample, 69% were White (non-Hispanic), 6% African-American (Black), 11% Asian-American (Asian), and 6% Hispanic-American. See Deming, G., & Hufnagel, B. 2001 "Who's taking ASTRO 101?" *Phys. Teacher*, 39, 368-369.

3. A meta-analysis (statistical review) of cooperative learning effects shows an effect size of about 0.5 in achievement attributed to small groups alone; see Springer, L., Stanne, M. E., & Donovan, S. S. 1999, "Effects of Small-group Learning on Undergraduates in Science, Mathematics, Engineering, and Technology: A Meta-analysis," *Review of Educational Research* 69, 21-51.

4. An on-line search of the American Journal of Physics using keywords "innovation", "transfer", "dissemination" individually and in combination with "astronomy" and physics" resulted in a maximum number of hits (26) for "innovation" by itself.

5. The attitude and concept map assessments are available at http://www.wcer.wisc.edu/nise/cl1; the current public release of the ADT (version 2) is located at http://www.aacc.cc.md.us/scibrhufnagel.

Appendix A: Brief on Educational Statistics

ANOVA

An analysis of the variance (ANOVA) is a test for statistical significance among means. ANOVA relies on the fact that variances can be divided up and attributed to random error and to the differences between means. In essence, ANOVA compares the variability between groups to that within groups and checks the latter for statistical significance. If significant to a standard probability level (p, usually 5% or less), then we can then accept the means as really different. All the statistical tests in this paper meet this standard ($p < 0.05$). ANOVA can handle many variables at a time.

t-test

If we are comparing only two means, we can use a *t-test* rather than ANOVA. 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 the means likely differ. Using 5%, we have a 95% chance of being correct. Note that the *p*-values for the t-test results in the tables are all 0.1% (or smaller). These are very robust results, from a statistics view.

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 the semester. It permits a calibration of comparisons across different characteristics of a study by normalizing the results by standard deviations. In the educational research, effect sizes of 0.1 or less are considered small and of no practical import; 0.3, medium and having practical significance, and 0.5 or greater, large (and unusual).

References

Bisard, W., & Zeilik, M. 1998, Conceptually Centered Astronomy with Actively Engaged Students, *Mercury*, 27, 16.

Bisard, W., Aaron, R., Francek, M., & Nelson, B. 1994, Assessing Selected Physical Science and Earth Science Misconceptions of Middle School Students through University Preservice Teachers, *Journal of College Science Teaching*, 24, 38.

Bonwell, C. C., & Eison, J. A. 1991, *Active Learning: Creating Excitement in the Classroom*, Washington, D.C.: The George Washington University, School of Education and Development.

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

Cummings, K., Marx, J., Thornton, R., & Kuhl, D. 1999, Evaluating Innovation in Studio Physics, *Phys. Educ. Res., Am. J. Phys. Suppl.*, 67, S38.

Fraknoi, A. 1996, The State of Astronomy Education in the United States, in *Astronomy Education: Current Developments, Future Coordination*, J. Percy (Editor), San Francisco: Astronomical Society of the Pacific, 89, 9.

Fraknoi, A. 2001, Enrollments in Astronomy 101 Courses: An Update, *Astronomy Education Review*, Issue #1.

Gouguenheim, L., McNally, D., & Percy, J. R. (Editors). 1998, *New Trends in Astronomy Teaching*, Cambridge University Press.

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, 64.

Halloun, I.A., & Hestenes, D. 1985, The Initial Knowledge State of College Physics Students, *American Journal of Physics*, 53, 1043.

Johnson, D. W., Johnson, R. T., & Smith, K.A. 1991, *Cooperative Learning: Increasing College Faculty Productivity*, Washington, D.C.: ASHE-ERIC Higher Education, Report No. 4.

Mazur, E. 1997, Peer Instruction: A User's Manual, Upper Saddle River, N.J.: Prentice-Hall.

McDermott, L. C. 1984, Research on Conceptual Understanding in Mechanics, *Physics Today*, 37, 24.

McDermott, L. C. 2001, Physics Education Research--The Key to Student Learning, *American Journal of Physics*, 69, 1127.

McDermott, L. C. & Redish, E. F. 1999, Resource Letter: PER-1: Physics Education Research, *American Journal of Physics*, 67, 755.

Meyers, C., & Jones, T. B. 1993, *Promoting Active Learning: Strategies for the College Classroom*, San Francisco: Jossey-Bass.

Michaelsen, L. K. 1992, Team Learning: A Comprehensive Approach for Harnessing the Power of Small Groups in Higher Education, in To Improve the Academy: Resources for Faculty, Instruction and Organizational Development, D. H. Wulff & J. D. Nyquist (Editors), Stillwater, Okla.: New Forums Press.

Michaelsen, L. K., Black, R. H., & Fink, L. D. 1996, What Every Faculty Developer Needs to Know About Learning Groups, in *To Improve the Academy: Resources for Faculty, Instructional and Organizational Development*, L. Richlin (Editor), Stillwater, Okla.: New Forums Press.

Millis, B. J., & Cotell, P. G., Jr. 1998, Cooperative Learning for Higher Education Faculty, Phoenix: Oryx Press.

Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (Editors). 1998, *Teaching Science for Understanding*, San Diego: Academic Press.

Novak, J. D., & Gowin, D. B. 1984, *Learning How to Learn*, Cambridge University Press.

Pasachoff, J. & Percy, J. 1990, *The Teaching of Astronomy: IAU Colloquium 105. Cambridge*, Cambridge University Press.

Redish, E. F. 1994, Implications of Cognitive Studies for Teaching Physics, *American Journal of Physics*, 62, 796.

Redish, E. F. 1999, Millikan Lecture 1998: Building a Science of Teaching Physics, *American Journal of Physics*, 67(7), 562.

Redish, E. F., & Steinberg, R. N. 1999, Teaching Physics: Figuring Out What Works, *Physics Today*, 52, 24.

Sadler, P.M. 1987, Misconceptions in Astronomy, in *Misconceptions and Educational Strategies in Science and Mathematics*, J. Novak (Editor), Ithaca, N.Y.: Cornell University Press, 422.

Sadler, P.M. 1992, The Initial Knowledge State of High School Astronomy Students, Cambridge: Harvard School of Education, unpublished dissertation.

Saul, J. M., & Redish, E. F. 1998, An Evaluation of the Workshop Physics Dissemination Project, preprint, University of Maryland Physics Education Research Group.

Schau, C., Mattern, N., & Teague, K. 1999, Select and Fill-In Concept Map Scores as a Measure of Students' Connected Understanding of Introductory Astronomy, paper presented at the annual meeting of the American Educational Research.

Van Heuvelen, A. 1991, Learning to Think Like a Physicist: A Review of Research-Based Instructional Strategies, *American Journal of Physics*, 59, 891.

Zeilik, M. 1996, Conceptual Astronomy: A Cognitive Approach for Teaching Science to Non-Majors, in *Astronomy Education: Current Developments, Future Coordination*, J. Percy (Editor), San Francisco: Astronomical Society of the Pacific, 295.

Zeilik, M., & Bisard, W. 1997, Conceptual Astronomy: A Cognitive Approach to Active Learning, paper presented at the meeting of the American Association of Physics Teachers, Phoenix.

Zeilik, M., & Bisard, W. 2000, Conceptual Change in Introductory-level Astronomy Courses, *Journal of College Science Teaching*, 29, 229.

Zeilik, M., Schau, C., & Mattern, N. 1988, Misconceptions and Their Change in University-Level Astronomy Courses, *The Physics Teacher*, 36, 104.

Zeilik, M., Schau, C., & Mattern, N. 1999, Conceptual Astronomy. II. Replicating Conceptual Gains, Probing Attitude Changes Across Three Semesters, *American Journal of Physics*, 67, 923.

Zeilik, M., Schau, C., Mattern, N., Hall, S., & Teague, K. W. 1995, Use of Concept Maps for Teaching and Assessing in Introductory Astronomy Classes, paper presented at the Rocky Mountain Education Coalition Conference.

Zeilik, M., Schau, C., Mattern, N., Hall, S., Teague, K. W., & Bisard, W. 1996, Conceptual Astronomy: A Model for Teaching Postsecondary Science Courses, paper presented at the meeting of the National Association for Research in Science Teaching, St. Louis.

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.

ÆR
33 - 46