

Astronomy Education Review

Volume 1, Oct 2001 - Jan 2002

Issue 1

Results from the Astronomy Diagnostic Test National Project

by **Grace L. Deming**
University of Maryland
Posted: 06/14/02

The Astronomy Education Review, Issue 1, Volume 1:52-57, 2002

© 2002, Grace Deming. Copyright assigned to the Association of Universities for Research in Astronomy, Inc.

Abstract

The Astronomy Diagnostic Test (ADT) is the first research-based assessment tool developed for use in undergraduate introductory astronomy classrooms. The ADT National Project has rigorously measured reliability and validity through the collection and analysis of a large sample of student ADT results.

The following was presented as part of the Special Session, "The Astronomy Diagnostic Test: Development, Results, and Applications" at the January 2002 Meeting of the American Astronomical Society.

1. INTRODUCTION

In 1998, a group of astronomers realized that in order to effectively evaluate and improve learning in undergraduate astronomy courses, a research-based assessment designed for introductory non-science majors was needed. During the next three years, astronomy education researchers (Collaboration for Astronomy Education Research) conducted student interviews, collected open-ended student responses to each of the test questions, solicited answers and critiques of the test from astronomy faculty, administered the revised versions of the Astronomy Diagnostic Test (ADT) as a pre-course and post-course test, and investigated reliability and validity of the assessment. In June 1999, the ADT Version 2.0 (ADT 2.0) was released (Hufnagel et al. 2000). (For a summary on the ADT's development, see Beth Hufnagel's paper, Development of the Astronomy Diagnostic Test, in this issue of *the AER*.)

In 2000, the National Science Foundation suggested that reliability and validity of the ADT 2.0 needed more rigorous examination. In educational research, reliability refers to the consistency of the information obtained by assessment; validity refers to the extent that the assessment provides the information desired (Wallen & Fraenkel 2001). Thus, for the ADT 2.0, a demonstration of acceptable consistency of student

responses in classes across the United States and at a variety of institutions would show reliability. In addition to demonstrating that a wrong answer usually means that a student does not understand a concept, there must be agreement on the correct answer among experts to show validity. An NSF Small Grant for Exploratory Research (REC-0089239) was awarded to examine the ADT 2.0's reliability and validity; the effort to formally investigate these quantities became known as the ADT National Project.

2. METHODS

Since reliability analyses require a large and representative sample of test results, the primary task of the ADT National Project involved the collection of student test results. A limited number of pre-course and post-course ADT 2.0 results had been collected from introductory astronomy classes offered during the fall 1999, spring 2000, and fall 2000 semesters. This collection of ADT 2.0 results included students in classes volunteered by interested colleagues, as well as students at the developers' home institutions. Detailed instructions for administering the ADT had been provided to instructors, and original student answer sheets were returned to researchers at the University of Maryland. During fall 2000, a major effort was planned for the spring 2001 semester that would increase the size of the sample.

Throughout the ADT's development, an ongoing effort was made to inform the community about our research (Hufnagel & Deming 1999). Careful records had been kept of contacts at professional meetings and inquiries about the ADT and other projects. A solicitation letter was sent to faculty on this contact list, inviting those with appropriate classes to participate in the ADT National Project. After responses were obtained, the list of participants was scanned for state and type of institution (university, four-year college, or two-year college). Since a representative sample was desired, additional solicitation letters were sent to astronomers in states not yet included in the sample and to institutions that had large classes. A final effort was mounted at the San Diego joint meeting of the American Astronomical Society and the American Association of Physics Teachers held during early January 2001 (Deming & Hufnagel 2000). Sixteen additional instructors volunteered their classes for the National Project and were given a packet on site. The packet provided to all participants contained a copy of the ADT 2.0 suitable for photocopying, scantron answer sheets, detailed instructions for administering the ADT, instructor and course profile data collection forms, and a mailer for returning completed student answer sheets to us. In exchange for participating, each instructor was sent a complete breakdown by question/choice of answer for each class, including gender values. Two weeks before the end of the course, a package containing post-course testing materials with instructions was sent to each participating instructor. Final class results were provided within a month after receipt of student responses, along with a chart containing anonymous sample statistics for different types of institutions, so that participants could compare their class results to others.

3. RESULTS

The participation of our colleagues teaching astronomy across the United States surpassed our expectations. The ADT National Project sample contains pre-course test results for 5,346 students and post-course results for 3,842 students. The sample includes 97 different classes that ranged in size from four to 320 students. The 68 participating professors taught classes in 31 states representing universities (50 classes), four-year colleges (25 classes), and two-year colleges (22 classes). The reliability analysis was performed at the Ontario Institute for Studies in Education (OISE). OISE calculated Cronbach's alpha values (Traub 1994) for both the pre-course sample and the post-course sample to estimate internal consistency (Gall et al. 1996). For the pre-course ADT 2.0 results, the value for Cronbach's alpha was 0.65. A value above 0.70 for Cronbach's alpha is desirable (Wallen & Fraenkel 2001). The slightly lower

pre-course value is indicative of some guessing and the high level of difficulty of ADT 2.0. Students taking their first college astronomy course see answers that mirror commonly held incorrect ideas, because the ADT developers based the selection of answers on student interviews. The pre-course test value for ADT 2.0 represents an improvement over the Cronbach's alpha value of 0.43 for a preliminary version of the ADT, ADT 1.0 (Zeilik et al. 1997). The post-course ADT 2.0 Cronbach's alpha value of 0.76 demonstrates an acceptable degree of internal consistency and reflects an improvement over the post-course ADT 1.0 value of 0.66 (Zeilik et al. 1997). Based on the reliability study and examining the individual class results from the ADT National Project, astronomy instructors can expect similar results if the ADT 2.0 were given in their introductory astronomy classrooms.

By considering the results from the student interviews, open-ended responses, and expert feedback, the validity of the ADT 2.0 was established. Forty-four of the 52 professors participating in the spring 2001 completed the ADT with an average score of 98% (standard error 0.5%). There was excellent agreement by the experts on the correct answer choices. Answers that did not agree with the majority were spread over several questions, rather than disagreement with a single question. As discussed in the development paper by Beth Hufnagel preceding this paper, student interviews provided confidence that students interpret questions and answers as expected by experts. When a student chooses a wrong answer, it usually means that he or she does not understand the concept.

The ADT national sample yielded an average value of 32.4% (standard error of 0.21%) for the pre-course test and 47.3% (standard error of 0.32%) for the post-course test. There is a gender discrepancy that persists in both the pre-course results (11% points) and the post-course (12% points) scores, although gains are similar (see Figure 1). There were no significant variations across geographic distribution, class sizes, or institution types.

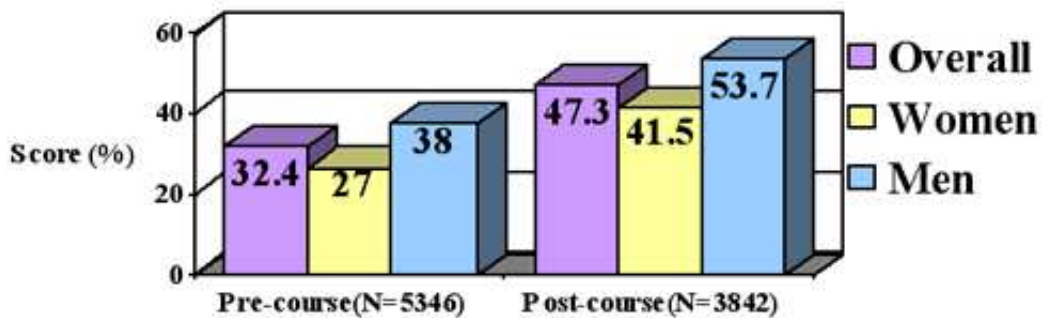


Figure 1. Pre- and post-course results for the ADT National Project

In addition to the 21 concept questions, the ADT 2.0 has 12 student background questions. We now know more about students who take a course in introductory astronomy, and a paper addressed the question "Who's Taking ASTRO 101?" (Deming & Hufnagel 2001) using a preliminary subset of the ADT National Project sample. When the entire pre-course sample was analyzed for 5,346 students, major choice and ethnicity of students taking introductory astronomy were very similar to all United States undergraduates (U.S. Department of Education 2000). In the U.S., more women than men currently attend college. However, the ADT National Project pre-course sample showed almost equal numbers of men and women taking introductory astronomy (see Figure 2). Women may be choosing a science other than

astronomy to satisfy general education requirements, and it would be an interesting project to investigate why our numbers don't reflect the overall U.S. numbers.

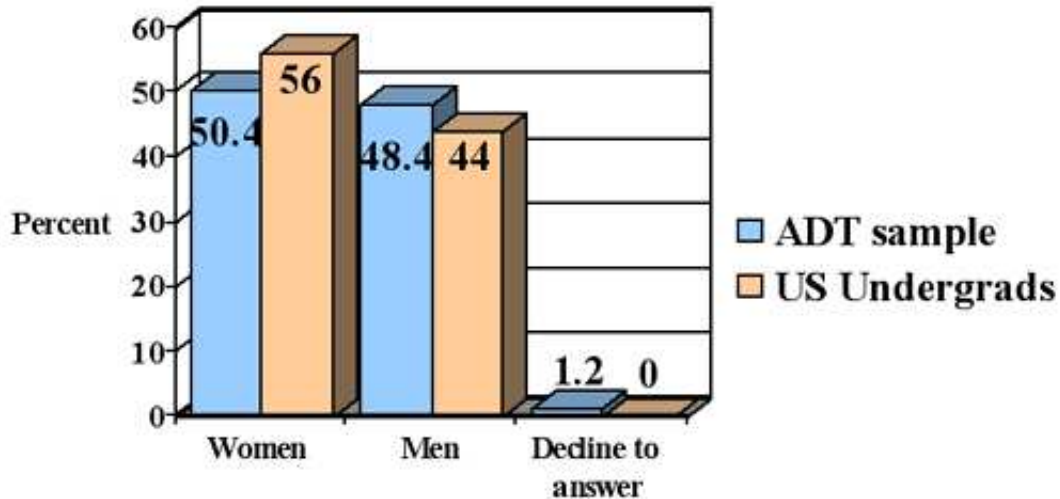


Figure 2. Gender mix of the ADT National Project sample

Students' confidence in math and science and their preparation in mathematics were also addressed in the background questions. Student confidence levels on the astronomy content questions on the ADT were low in the pre-course national sample. There was a discrepancy between confidence levels expressed by women and men. Although there was some improvement in confidence in the post-course national sample, the gender discrepancy persisted (see Figure 3). Investigating the gains in confidence levels over a variety of teaching methods merits additional study. When asked about their last math course completed, the majority of students in the pre-course sample indicated that they had taken calculus (34%) or pre-calculus (22%). The remaining students had completed algebra (28%), geometry (5%), trigonometry (9%), or didn't answer the question (2%).

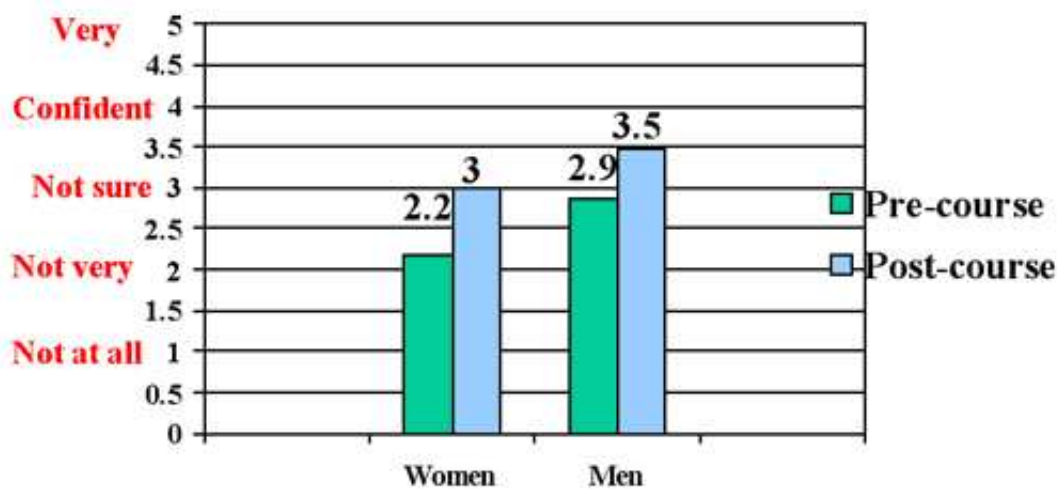


Figure 3. Student confidence in ADT answers

4. SUMMARY

The ADT 2.0 is a reliable and valid assessment of student ideas concerning a limited number of astronomy concepts. Pre-course testing provides instructors with information about misconceptions held by their students and can indicate the diversity of their students. Awareness of the prevalence of incorrect student ideas in the college classroom permits instructors to focus on problem areas and select teaching methods tailored to each class. Post-course testing can provide an indication of student learning. The fairly low post-course scores permit researchers to use the ADT for evaluation and comparison of teaching methods. With the ADT 2.0, astronomy instructors now have a reliable and valid tool for understanding student ideas prior to instruction and how those ideas change after completion of an introductory astronomy course.

The author would like to thank all those who participated in the Astronomy Diagnostic Test National Project, and undergraduates Elizabeth Miller and Krista Snyder, who assisted in data reduction and correspondence with participants. This work was supported in part by National Science Foundation grant REC-0089239.

References

- Deming, G. & Hufnagel, B. 2001, Who's Taking ASTRO 101?, *The Physics Teacher*, 39(6), 368.
- Deming, G., & Hufnagel, B. 2000, The Astronomy Diagnostic Test: Past, Present, and Future, *BAAS*, 32(4), 1154.
- Gall, M., Borg, W., & Gall, J. 1996, *Educational Research*, White Plains, N.Y.: Longman Publishers.
- Hufnagel, B., & Deming, G. 1999, The Astronomy Diagnostic Test: Comparing Your Class to Others, *BAAS*, 31(3), 937.

Hufnagel, B., Slater, T., Deming, G., Adams, J., Lindell Adrian, R., Brick, C., & Zeilik, M. 2000, Pre-course Results from the Astronomy Diagnostic Test, *Publications of the Astronomical Society of Australia*, 17:2.

Traub, Ross E. 1994, *Reliability for the Social Sciences: Theory and Applications*, Thousand Oaks, Calif.: Sage Publications.

U.S. Department of Education, National Center for Education Statistics. 2000, Fall Enrollment in Colleges and Universities Surveys.

Wallen, N. & Fraenkel, J. 2001, *Educational Research: A Guide to the Process*, Mahwah, N.J.: Lawrence Erlbaum Associates.

Zeilik, M., Schau, C., Mattern, N., Hall, S., Teague, K. W., & Bisard, W. 1997, Conceptual Astronomy: A Novel Approach for Teaching Postsecondary Science Courses, *American Journal of Physics*, 65(10), 987.

ÆR

52 - 57