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

Astronomy Education Review

Volume 4, Oct 2005 - Jul 2006
Issue 2

Promoting Undergraduate Critical Thinking in Astro 101 Lab Exercises

by **Michael L. Allen**
Washington State University Department of Physics & Astronomy
Diane Kelly-Riley
Washington State University Critical Thinking Project and Writing Program
Received: 08/30/05, Revised: 10/03/05, Posted: 11/04/05

The Astronomy Education Review, Issue 2, Volume 4:10-19, 2006

Abstract

This article presents results of the first two years of the introduction of a critical thinking (CT) component to standard freshman astronomy lab exercises for nonmajors. The component consists of a series of probing questions folded into the exercises, plus a formal grading rubric. The grading rubric is adapted from the generalized Washington State University *Guide to Rating Critical Thinking* developed by the Washington State University Critical Thinking Project. The questions asked of the students are very specific applications of the more general *Guide*. The motivations, implementation, and results of the use of this CT component are presented, as well as some ideas for further improvement.

1. INTRODUCTION

Many undergraduate astronomy courses feature a lab component. This laboratory component benefits students in many ways: students learn better when they are actively engaged with the material (e.g., Slater & Adams 2003), they are exposed to the collection and treatment of data, material presented in the lecture component of the course is reinforced, and students working in groups learn better based on theories of learning as a social activity (e.g., Green 2003).

There are many sources for lab exercises, the choice of which depends upon the preferences of the instructor and the resources available in the lab. These sources include periodicals (e.g., *Sky and Telescope* magazine), published collections (e.g., Nicastro 1990), standalone computer exercises (e.g., Project CLEA), Internet-based data mining (e.g., SDSS SkyServer), and observing projects with small telescopes (e.g., Gainer 1974).

Despite the wealth of resources and the manifest benefits of lab exercises, there is still some question concerning the depth of understanding achieved by the students. Physics-specific education studies (for a review, see McDermott & Redish 1999 and Bailey & Slater 2005) and a popular video series (*A Private Universe* and *Minds of Their Own*) find that simple phenomena, such as the cause of the phasing of the Moon, are misunderstood by a potentially large fraction of the student population, and by inference, the general population.

Because the performance of lab exercises appears to be a step toward correcting these misunderstandings, one can ask how the lab experience can be enhanced to best communicate the material to students. Teachers must cope with the tension between communicating underlying physical principles and the application of these principles. Furthermore, teachers hope that their students will be able to understand the results generated. A resolution to this tension comes from looking at the lab exercises themselves, and the students' performing of them.

Typical lab exercises, as they currently exist, are mostly cookbooklike series of steps that, when followed, generate some combination of numbers and diagrams. The exercises do not require students to develop a deeper understanding of the course material or encourage them to apply their knowledge.

To illustrate the tension between content and application, the instructor (MLA) recalls an episode in a lab period with a group of approximately 20 students. Two of the students were working together very diligently, punching buttons on a calculator, comparing their notes with the lab instruction sheet, and scribbling down numbers. Finally, they sprang up from their seats, shouting, "We did it!" They clasped hands and for a moment were quite animated. Then one of the pair asked, "What did we just do?" to which the other replied, "I don't know."

Such an exhibition demonstrates a prevailing problem with undergraduate lab exercises: that students mechanically proceed through the procedure without knowing *why* they are doing what they are doing, *how* the instructions lead to the final result, or *what* that final result is going to be (the colloquial "big picture"). Clearly, the quality of students' lab experiences must be improved. At the time of the observation of the phenomenon of students arriving at the correct answer without knowing why, Washington State University was implementing a campuswide Critical Thinking Project (the Project) based on the locally developed WSU *Guide to Rating Critical Thinking* (the *Guide*; see Appendix B). The Project encouraged faculty to adapt a seven-dimension rubric to articulate their instructional and evaluative expectations for critical thinking (CT). Faculty across the disciplines were encouraged to operationalize their definitions of critical thinking based on the conventions of their disciplines, the size of their courses, and their personal teaching styles (Kelly-Riley, 2003).

The astronomy undergraduate lab component for nonmajors came together with the Critical Thinking Project in the fall of 2002. Although this article discusses labs for nonmajors, the methods and ideas are applicable to lab exercises at all levels and within the major as well.

This article discusses observations from the first two years of use of these revised exercises adapted from the WSU *Guide*. The *Guide* serves as a way of translating, interpreting, or judging a student's performance in comparison with preset expectations. The motivation for integrating CT expectations into undergraduate labs was presented in section 1. Section 2 discusses the Washington State University Critical Thinking Project; section 3 examines the lab exercises and their modification; section 4 discusses the theory behind the adaptation of the *Guide*; and section 5 provides a personal account of the benefits realized and the remaining problems to be solved by the lab instructor, plus suggestions for further work.

2. THE WSU CRITICAL THINKING PROJECT

The Washington State University Critical Thinking Project was a campuswide effort to promote instructional and evaluative strategies that encouraged student critical thinking in undergraduate classrooms. The Project received funding from the Washington State Higher Education Coordinating Board and the U.S. Department of Education Fund for the Improvement of Postsecondary Education (FIPSE) comprehensive program to combine assessment with instruction. The goal of this effort was to increase coherence and promote higher-order thinking and to encourage improvement of faculty teaching and evaluative practices in a four-year general education curriculum at a large Research I public university.

The Project developed the seven-dimension Washington State University *Guide to Rating Critical Thinking* (http://wsuctproject.wsu.edu; see Appendix B) to provide a path to improve student learning, reform faculty teaching, and provide a means to measure the effectiveness of teaching and learning at our institution. The Project provided ongoing support for faculty to implement innovative combinations of teaching and assessment methodologies in the three tiers of general education courses that span all the disciplines. Faculty were encouraged to operationalize their definitions of CT and were not told to follow a single definition or application of CT (Condon & Kelly- Riley, 2004). There was no obligation to apply each of the seven dimensions, and it was perfectly acceptable to make the introduction of CT expectations slowly, over many years.

Samples of student papers were assessed using a six-point scale outside of the regular classroom by faculty from across the disciplines; they used a rating methodology developed by the Project (Condon & Kelly-Riley, 2004). All student papers that were exposed to the methodologies in the WSU Critical Thinking Project demonstrated a statistically significant higher difference than their undergraduate counterparts who were not exposed to the CT expectations.

3. ADAPTING THE LAB EXERCISES TO PROMOTE CT

In general, any lab exercise should (1) be easily performable by the students, (2) be understandable on a step-by-step basis within the context of the goal, and (3) have a meaningful goal. Our freshman-level lab exercises are of the pencil-and-paper type, where students are given precollected data to analyze. A few examples of these exercises include asking students to design a scale model of the Solar System, reconstruct the orbit of a planet given some observations, draw conclusions about the properties of stars based on a survey of stellar properties, and use a data set to demonstrate that there are planets around other stars. These exercises are practical in nature, generate a numerical result, can be verified against numbers in the appendixes in a textbook, and mimic the formal experiments performed by professional astronomers. In addition to the more procedural and routine exercises, there are two open-ended exercises. In one, the students are asked to determine how well the eye can distinguish two closely placed objects but are not given a formal procedure. In another, the students are asked to design a space mission. These latter exercises are great favorites.

In the first semester, to use the grading rubric (Appendix A), the lab exercises were left unrevised from their original sources. The rubric had a number of general questions that the students were asked to address in their lab reports. A few examples of these questions are:

- How does the procedure lead directly to the final result?
- What assumptions allow the lab procedure to work?
- Are the conclusions supported by the data?
- How do your results compare with the accepted values?

This approach did not aid the students. Students were unable to apply a general question to a specific problem. For example, the students could not identify even the most basic assumption in an exercise; they had to be told what the assumptions were.

Based on this first experience, it was obvious that the lab instruction sheets had to be rewritten, with the critical thinking aspects framed as specific, and often leading, questions. Whereas in the original exercise, the students were asked to identify an assumption, in the new exercises, the students were asked to evaluate the importance of a specific assumption. When determining the orbit of Mercury, for example, one must start by drawing the orbit of the Earth. The students were told to assume a circular orbit for the Earth and then were asked, "The Earth's orbit is not circular, so how big an error are you introducing into the exercise?" The answer is an error of less than 2%. A follow-up question asks, "Is Mercury's orbit unquestionably noncircular?" The conclusion that the student should come to, in the absence of formal error analysis, is that Mercury's orbit is highly noncircular, so even a 2% change will not circularize Mercury's orbit.

Each lab exercise ends with a set of questions to which the students must submit written answers. Each of these questions falls under one of the 10 categories in the grading rubric (Appendix A). Therefore, the questions have a one-to-one correspondence with the grading rubric. For example, when the students are writing down an answer to a question categorized as a procedure-type question, they know that the grade for their answer is judged against the procedure-type behavior anchors described in the grading rubric. It is possible that more than one question will apply to a single criterion.

How were these questions formulated? How do they fit into the framework of the lab exercises and promote CT? The general questions in the grading rubric, as adapted from the *Guide*, were used as the starting point.

The WSU CT Project identifies seven CT activities. Not all seven apply to each lab exercise. For example, the student's own perspective (criterion 2 in the *Guide*) is not asked for. Perspective in science does not refer to agreeing with the accepted answer. Rather, it refers to an interpretation of the accepted answer and the method used to obtain it. For example, one can interpret gravity in a Newtonian perspective (action at a distance) or an Einsteinian perspective (curved space-time). Both interpretations can yield predictions verifiable by experiment, but both appeal to a different mechanism for generating gravity. This level of comparison is beyond freshman non-science majors, so it is not asked of them.

The other six CT criteria in the *Guide* are addressed by conscious choice of the instructor. In the first draft of the exercises, which was tested for two semesters, there was a question for each of the remaining six CT criteria. The result is that some of the questions in the lab exercises were extremely difficult to answer, even for an expert.

For example, a difficult question to answer asked in every lab exercise during these first two semesters was, "Write a brief summary of the procedure (omit technical details). Do you believe that the goal was achieved? Why?" The idea behind this question was to encourage students to move away from thinking of the exercises as a set of instructions to be blindly followed and instead to question why the exercise was performed in a particular way.

From reading student papers, it was clear that the students did nothing more than write out the procedure like a cookbook; there was little or no attempt to communicate a deeper understanding. It could not be told from their answers whether they understood the procedure; it was clear, however, that the spirit of the question was not addressed by the students. In the next year (2004–2005) of using the exercises, this particular question had been reinvented to ask about one or two crucial parts of the procedure and why the lab exercise would not work without them.

The same grading rubric was used to judge each exercise. Use of the same rubric allowed both the student and the grader to become familiar with the level of performance expected of them. The grading rubric is, therefore, general in nature. Perhaps the level of performance expected of the student would be better communicated if a different lab rubric was constructed for each exercise.

4. ADAPTING THE *GUIDE*: A HOW-TO APPROACH

The *Guide* was adapted to the freshman-level course, Descriptive Astronomy, a survey course for nonmajors. The method, which was individual to the instructor, was painfully sequential in nature but produced a usable grading rubric for students and instructors. The grading rubric was constructed first and then used as a guide for constructing more specific questions to be asked of the students at the end of each lab exercise.

The method followed by the instructor here was:

1. Compose a philosophy for a general rubric pertaining to the goals of the lab section.
2. Brainstorm as many grading criteria as possible.
3. Arrange criteria into logical groups.
4. Consolidate and prune.
5. Refer back to the *Guide*.
6. Test and revise the rubric based on results from its use.

The instructor’s philosophy was straightforward but in some cases highly individual. The following principles were used when trying to reconcile the tensions between content and critical thinking:

1. Teach your subject first, CT second.
2. Grade on both knowledge and CT ability.
3. Every assignment should promote critical thinking at some level.
4. The grading rubric must be enlightening to the student.
5. The grading rubric must be usable and lighten the workload of the graders.

4.1 Teach Your Subject First

At the postsecondary level, instructors are well-versed in their subject areas but not necessarily in effective teaching methods. It is probably best for instructors to play to their strengths and devote much of their time to helping students acquire and organize new knowledge. CT assignments and evaluation criteria can be vehicles for teaching and learning new subject matter.

4.2 Grade on Both Knowledge and CT Ability

The ability to understand a process or describe some situation perfectly is valuable. Every student takes a different amount of time to feel comfortable with new knowledge. Regrettably, our society's stopwatch style of education hinders those who need more than a few weeks or months to think about and completely understand an idea. The only purpose of a semester is to make the process of education convenient for teachers of large numbers of students. There should not be a penalty for a learn-by-rote student who may gain a fuller understanding of the material when reflecting on the material at a later time. Therefore, the lab rubric has two parts: one to complete the lab exercise and one to encourage students to reflect on the exercise itself.

4.3 Every Assignment Has a Critical Aspect

Before implementing a CT component, the exercises were of the type in which the student would plot the data, extract a trend, and thus see a demonstration of some well-known result. Now the exercises are turned around so that the student is given a problem to solve and a method for solving it, and is asked if he or she believes the result and at what level this belief (or disbelief) is justifiable. The seven aspects of CT as outlined in the *Guide* are very general and must be made specific to any lab investigation. The word *investigation* is important. Each CT lab exercise begins with a question to be answered. Then, one can further ask, for example, "is the method relevant?" and "what is the certainty of the answer?"

4.4 The Rubric Must Be Enlightening to the Student

The lab rubric, which is identical for every lab exercise, attempts to appeal to the student by giving sets of three or more behavior anchors for each grading criterion. In other words, there is a written description of the level of proficiency required to obtain a certain minimum grade (see Appendix A). Each of these criteria is rated on a five-point scale, divided into three tiers. A low level of performance earns 0 or 1 of 5 total points, a medium level earns 2 or 3, and a high level earns 4 or 5; these levels have descriptions of what low, medium, and high mean in practice. For example, when describing the procedure of an exercise, a low performance level is described as "copied from the work sheets," a medium level as "paraphrased cookbook format," and a high level as "paraphrased as a means to an end."

The grading rubric translates directly into a numerical grade. The decision to commit the lab rubric to generating a numerical grade was made in this case by the instructor, based upon the personal belief that a grading scheme is simply not a proper one in the absence of some way for the students to know how well or poorly they are doing, even if in a relative sense only.

The act of constructing and using a detailed grading rubric is useful in that many instructors find that their teaching methods change. For example, the emphasis on some material may change to prevent it from being either (1) insignificant in the rubric or (2) dominant. We all recognize a good paper when we read it; to articulate what constitutes a good paper is a difficult but necessary task.

4.5 The Rubric Must Be Useful for the Graders

The rubric must accurately reflect the grader's opinion of the student's work. The rubric must also be quick to use, because a typical enrollment for the undergraduate survey course is 150 students. Ease of use for the grader includes the direct translation to a numerical grade, about which enough has been said. The original lab rubric had 12 criteria. After two semesters of use, this number was reduced to 10 without a significant loss of integrity.

5. RESULTS AND FUTURE WORK

During the years 2003–2005, the efficacy of the CT component of the labs was judged in two ways, one informal, and one more formal. Informally, the instructor talks regularly with both the students and graders and performs a survey to discover the students' most- and least-liked labs. More formally, the WSU CT Project generates a biannual report from a set of independent graders about the students' CT abilities and how the lab exercises address these abilities. This report is automatically given to those connected with WSU's CT Project.

5.1 Informal Results

In the first course for which a CT lab rubric was used, the lab grades increased, but so did the standard deviation. Looking at the grade distribution, there are at least two sources for this increase in the variance of scores: (1) some students do not adapt over the course of one semester and slip behind their peers (some of these students did not take full advantage of the benefits of peer interaction), and (2) the students become tired and put less effort into their lab reports. The number of late and missing lab reports increases with time. Each lab report is worth about 2% of the final grade, so the perceived consequences of missing one or two are slight.

Do the students improve their CT skills? In speaking with many students, there was a detectable change in the manner in which they expressed themselves when asking questions in the lab. The students more readily used scientific terminology (i.e., jargon), and the content of their questions became less general. In reading lab reports, there were some areas in which the students learned to express themselves better or learned what type of argument was expected of them. For example, the idea of comparing an expected error with a measured error (both numerical values) has become routine. An important example of improvement harkens back to an early and primary concern of the instructor—namely, the students became demonstrably more aware of the goal of the exercises than they were in the past.

5.2 Formal Results

The formal results are given here for the sake of completeness. These results are difficult to interpret because the lab exercises were changing while the data was being collected. The changes, described in previous sections, were motivated by the desire to improve the lab exercises based upon the feedback received. Additionally, the group of faculty acting as reviewers (see below) was also changing during the process.

Here is how the formal results were generated: each semester, a group of faculty involved in the WSU CT Project would review (1) the lab exercises and (2) a selection of student lab reports chosen at random from the class and representing about 8% of the total number of students. The reviewers would rate each assignment and student report in each of the seven different CT categories in the *Guide*. The ratings were on a six-point scale. For rating the assignment, the reviewer was to ask, "To what extent is each aspect of CT asked for?" For student reports, each reviewer was to ask, "To what extent is each aspect of CT present?" Scores of 1 or 2 indicated that that aspect of CT was relatively absent. Scores of 3 or 4 indicated that that aspect of CT was only implicitly present. Scores of 5–6 indicated that that aspect of CT was explicitly present.

In Table 1 below, only descriptive ratings are given. When two descriptors are given, it means that the numerical value of the faculty rating was on the border between two categories. The class of fall 2003 was surveyed twice, once early in the semester (column 3) and once late (column 4).

Table 1. Descriptive Ratings

"To what extent is critical thinking present in student lab reports?"			
CT Dimension	**Spring 03**	**Fall 03 (1)**	**Fall 03 (2)**
1. Problem/question	Implicit	Explicit	*Implicit*
2. Student position	Absent/Implicit	Explicit	Implicit
3. Other positions	Implicit	Absent/Implicit	Absent
4. Assumptions	Absent/Implicit	Explicit	Absent
5. Evidence quality	Implicit	Explicit	Implicit
6. Context	Absent/Implicit	Implicit	Implicit
7. Implications	Absent/Implicit	Implicit	Absent/Implicit
"To what extent does the assignment evoke critical thinking?"			
CT Dimension	**Spring 03**	**Fall 03 (1)**	**Fall 03 (2)**
1. Problem/question	Implicit	Implicit	*Explicit*
2. Student position	Absent	Explicit	Implicit
3. Other positions	Absent	Explicit	Implicit
4. Assumptions	Absent/Implicit	Absent	Implicit
5. Evidence quality	Implicit	Explicit	Implicit
6. Context	Absent/Implicit	Implicit	Implicit
7. Implications	Absent/Implicit	Explicit	Explicit

The results in Table 1 are heavily subjective. However, a few trends can be seen. For example, by comparing the two tables, one can estimate to what extent students respond to the exercise. Consider the second evaluation of the class of fall 2003 (column 4). The italicized entry compares the students' ability to think critically (top table) with the assignment's attempt to evoke critical thinking (bottom table) for CT dimension 1 (that students are expected to know the overall goal of the exercise). In the fall 2003 semester, the reviewers thought that this knowledge was explicitly asked of the students, yet the reviewers also thought that the students were only responding to this requirement in an implicit way.

5.3 Future Work

In the future, it would be desirable to add a few more lab exercises, with emphasis on exercises in which the students can collect their own data; in many of the current labs, students are handed complete or partial data sets for analysis. Collecting their own data would help the students better understand the procedure of collection, and measurement uncertainties. Also, students enjoy data collection, and this enjoyment motivates the students to put more intellectual effort into the lab exercise.

Another component that should be added is an in-lab pretest and posttest, with conceptual questions not tied closely to the course content. Astronomy-based pre- and posttests are available from a number of sources (e.g., the Astronomy Diagnostic Test of the CAER) yet do not test problem-solving ability or critical thinking. The type of testing proposed here will be crucial to the further development of CT abilities in the students. Currently, the only data collected are informal: the instructor talks regularly with both the students and graders and performs a survey to discover the students' most- and least-liked labs. There is a great opportunity for us as astronomy educators to think about ways in which we can promote the higher-order learning abilities of students. Experimentation and collaboration are effective ways for us to operationalize our definitions of critical thinking and will help articulate our expectations to students so that they can answer the question of how they arrived at a correct solution.

Any reader wishing to have a copy of the lab exercises developed and discussed here can contact the authors at mlfa@wsu.edu.

Acknowledgments

Michael L. Allen thanks the members of the Critical Thinking Project for making their research available for use in a practical way and for hosting an interesting series of seminars and peer discussions, all of which have been useful for forming and articulating the ideas presented here. The WSU CT Project is supported by a grant from the Fund for the Improvement of Postsecondary Education (FIPSE).

Printed References

Bailey, J. M, & Slater, T. F., 2005, *American Journal of Physics*, 73, 677.

Condon, W., & Kelly-Riley, D. O., 2004, "Assessing and Teaching What We Value: The Relationship between College-Level Writing and Critical Thinking Abilities," *Assessing Writing*, 9, 56-75.

Gainer, M. K., *Astronomy: Observational Activities and Experiments*, Boston: Allyn & Bacon, 1974.

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

Kelly-Riley, D. O., 2003, "Washington State University Critical Thinking Project: Improving Student Learning Outcomes through Faculty Practice," *Assessment Update*, 15(4), 5.

"Laboratory Exercises in Astronomy." A series of articles in *Sky and Telescope* magazine, Cambridge, MA: Sky Publishing.