

# Assessing Student Learning in the Responsible Conduct of Research<sup>1</sup>

*Kenneth D. Pimple, Ph.D.*

April 2001

## Introduction

As with all effective teaching, a key element in any unit designed to teach research ethics is assessment of student learning, whether the “unit” is a whole course, a brown bag lunch seminar, or a few sessions or segments of sections within a course. How can we know what our students learned, and whether they learned what we intended them to learn? How can we tell whether our instructional goals were met?

I am reminded of the Saturday Night Live skit in which a reporter asks President Carter, “Why are you building the B-1 Bomber?” To which Carter replies, “I’m going to take that question in two parts. First, ‘Why?’ This is a question that has baffled philosophers and religious leaders for millennia, and it’s hardly fair to expect me to answer it during a press conference. Second, ‘Are you building the B-1 Bomber?’ Yes. Next question.”

I am addressing a different question: “How can we assess student learning in research ethics?”

To take it in two parts: “How?” This is obviously a tremendous question that scientists and philosophers have invested countless hours in studying and I can’t possibly begin to address it in one short paper. Second: “Can we evaluate educational programs or assess student learning in research ethics.” Yes.

That isn’t all that helpful, no matter how true. So let me divide the question a different way.

Part I: “How can we assess student learning (full stop)?” This time I am not joking when I say that this is a huge question that educators have addressed for centuries, and it would be

---

<sup>1</sup> Adapted from a presentation at the Planning Workshop for a Guide for Teaching Responsible Science, sponsored by the National Academy of Sciences, the National Science Foundation, and the National Institutes of Health, February 1997.

Copyright 2001, Kenneth D. Pimple, Ph.D. All rights reserved.

For information about this work, please contact Kenneth D. Pimple, Ph.D., Director of Teaching Research Ethics Programs, Poynter Center for the Study of Ethics and American Institutions, Indiana University, Bloomington IN 47405-3602; (812) 855-0261; FAX 855-3315; [pimple@indiana.edu](mailto:pimple@indiana.edu); <http://poynter.indiana.edu/>; <http://mypage.iu.edu/~pimple/>.

Permission is hereby granted to reproduce and distribute copies of this work for nonprofit educational purposes, provided that copies are distributed at or below cost, and that the author, source, and copyright notice are included on each copy. This permission is in addition to rights of reproduction granted under Sections 107, 108, and other provisions of the U.S. Copyright Act. Before making any distribution of this work, please ascertain whether you have the current version by contacting the author or the Poynter Center or checking <http://poynter.indiana.edu/tre/>; follow the link to “Resources for teaching research ethics,” and then to “Assessing Student Learning . . . .”

impossible to go over even a fraction of it here. A great deal is known about assessing student learning, and many techniques that can be used to assess student learning in any field can be used here as well, when suitably adapted. For example, if one of your goals is to teach objective information, such as your institution's regulations and policies on animal use, you can use any number of methods to test whether students have acquired that knowledge, including all of the methods you normally use when you assign grades.

But then we get to the second part of the question, "How do you assess student learning in research ethics?" We have to accept that some of the goals of teaching responsible research are different from many other kinds of teaching. But we also have to understand that teaching research ethics is not utterly unique; it has many points of similarity with other kinds of teaching.

## Special challenges in assessing learning in ethics

Persons who have not taught ethics often balk at the thought of grading students in an ethics course or students' performance in an ethics module. There are several possible reasons for this. First is the long reign of positivism and the virtual banishment of discussion of morality and ethics from higher education in the United States for most of this century. In the positivist paradigm, values could not be measured or weighed, and therefore values were seen as merely a matter of opinion, not a proper subject for instruction or assessment.

Second is the influence of pluralism in the United States. We are proud of the fact that no one can force religious observance or belief on anyone else. But a belief in religious tolerance can sometimes be overextended and lead to doubts about whether we share any values. While there are certainly areas of disagreement – and some of them capture headlines daily – I think it is also clear that there are many, many values that most Americans share and endorse.

Perhaps more subtly, the very importance and potency of ethics and morality in everyday life leads to doubt about whether it can be taught and whether the teaching of ethics can be assessed. We rightly recognize that ethics is an especially important domain of human belief, experience, and behavior. We also tend to think of morality as a central part of the self; in this it is similar to creativity or intelligence in that we think of it as a characteristic that one has rather than as a skill that one learns (like doing quadratic equations or operating an electron microscope) or as knowledge that one acquires (like the Pythagorean theorem or the atomic weight of helium).

Typically, we do not hesitate to grade our students based on their skills or knowledge, but we do not really want to – and, I suggest, we should not – grade them on their character.

But teaching research ethics is not just about character. It is also about skills, like moral reasoning, and knowledge, like rules and regulations. We would probably be happy if our students left our research ethics courses morally better than when they came in; if their character improved. But we don't have to assess their character. Assessing relevant skills and knowledge is sufficient.

## Unrealistic expectations

Unfortunately, this importance of morality can translate into unrealistically high expectations for instruction in ethics. People ask, "Do you really think you can make people ethical by making them take a course?" My answer is no. I do not expect to convert anyone. I do not expect my students who are cheaters and plagiarists to see the light and mend their ways after one

course. If they don't, does that make my course a failure?

Well, we do not expect students who are bad writers and have no imagination to become great novelists after one course in creative writing, or students who are nearly innumerate to become talented mathematicians after one course in algebra, and we do not thereby conclude that those courses are failures.

Students do not become novelists, or mathematicians, or ethical researchers, by taking one course. But they can make progress toward becoming novelists, mathematicians, or ethical researchers in just one course if they have the raw material for being so, and if they are willing to work at it.

It should be obvious that you can never make a novelist out of someone who doesn't want to be a novelist and isn't willing to work at being a novelist. But a good teacher can help even reticent students become better at writing. The same holds true in the realm of research ethics. No one who is basically dishonest is likely to be changed by education in the responsible conduct of research. But a good teacher can help students become better at recognizing ethical problems and thinking through solutions – even those students who begin by thinking that a course on research ethics is a waste of time.

## Realistic goals

As with any kind of assessment, our first goal has to be to clarify what we want to teach. It is unrealistic to suppose that we can assess whether we have made our students into ethical researchers. It is also unrealistic to suppose that we can assess the long-term impact of our teaching. It would be wonderful to know this, but do we know the long-term impact of any particular course?

Deni Elliott and Judy Stern make an interesting and useful distinction between “pedagogical hopes” and “instructional objectives.” As they put it,

One might have pedagogical hope that one's students become highly ethical practitioners in their careers and become highly ethical people in their private and public lives as well. But, that is not an objective that can guide the teaching plan for a specific class period. [Elliott and Stern 1996:346]

Another way to make the distinction is to think about proximate and ultimate goals. One proximate goal may be improving our students' moral reasoning ability. Ultimate goals may be training a generation of highly responsible researchers, or reducing the incidence of misconduct in research. But how do you assess your success toward the latter goals? That has to be left to social scientists and future historians. No given educator can be expected to answer questions like these about her or his own classes.

Whenever we want to teach something, we should be clear on what we want to teach; and whenever we want to assess our teaching, we have to be clear on exactly what we expect to be able to assess. We also have to remember that assessment comes in two flavors: quantitative and qualitative. Many of us have a bias toward numbers; if we can't quantify results, they don't count. That is a short-sighted view and eliminates many good methods for assessing student learning; it also might make it seem impossible to assess student learning in ethics. After all, how can ethics be reduced to numbers?

## Portfolios

Penny Gilmer,<sup>2</sup> a chemist at Florida State University, has successfully adapted a qualitative approach often used in education and creative arts programs to teaching research ethics: The use of portfolios.

Portfolios work like this: Students have a number of assignments through the semester, for example, writing essays, responding to case studies, collecting relevant clippings from the popular press, keeping a journal, and the like. At the end of the semester (or perhaps more often), they assemble their best efforts in a portfolio, which they submit for grading. One obvious strength of this method is that it allows students to compensate for their weaknesses by showcasing their true talents. For example, a student who has a hard time writing an analytical response to a case study might excel at spotting ethical issues in the press, or in some other area.

## Pre-test, post-test

In “Evaluating Teaching and Students’ Learning of Academic Research Ethics,” Elliott and Stern (1996) describe their initial, unsuccessful attempt at evaluating student learning in a graduate-level seminar in academic research ethics at Dartmouth College, as well as the use of a second strategy to evaluate student learning, which they deemed more successful. The course was developed by a team of faculty members and supported by grants from the Department of Education’s Fund for the Improvement of Postsecondary Education (FIPSE) and by the National Science Foundation.

In evaluating the success of the course, the instructors tried to assess both what the students learned (in terms of both skills and content) and the learning environment of the classroom. The latter is important because

Learning applied ethics requires students to take intellectual risks; it requires students to give close examination to their beliefs, values and methods of thinking about adequate and inadequate professional behavior. Clearly, some environments encourage this kind of risk taking and other environments discourage it. [352-353]

The team’s attempts to measure the learning environment included development of a fairly elaborate instrument, which was ultimately deemed unsuccessful; they concluded that the use of a standard student evaluation is adequate for this purpose.

As Elliott and Stern point out, a prerequisite for evaluating the success of a course is a clear understanding of the instructional objectives of the course, and their goals are particularly well-stated. At the end of the course, students will

- 1) be able to clearly describe relevant scientific conventions including: laboratory practice, institutional responsibility, etc.;
- 2) be able to describe what leads to ethical problems including causes inherent in the social context of the practice of science;
- 3) be able to identify ideal scientific practice and consider how to bring scientific conventions more in line with the ideal;
- 4) be able to separate behaviors into four categories: morally prohibited, required, permitted, and

---

<sup>2</sup> Dr. Penny J. Gilmer, Department of Chemistry, Florida State University, Tallahassee FL 32306; 850-644-4026; gilmer@sb.fsu.edu.

encouraged, thus illustrating an understanding of the role of the scientist in society. [349]

To measure the success of the course, the team of instructors used a pre- and post-test method. Extensive efforts were taken to assure the validity and reliability of the evaluation, including the use of three outside scorers and a fourth outsider who compiled and analyzed the results. The test required students to read an edited article on an actual case of misconduct of science taken from the journal Science (included as an appendix to the article) and answer the following question:

Identify the ethics problems in this case. Discuss what the individuals involved did right. Discuss what the individuals involved could have or should have done differently. [350]

The results were disappointing. There was no inter-rater reliability, and “there was no significant difference between how students approached the vignette at the beginning of the term and how they approached it at the end.” The authors identify several reasons for this failure.

We realized that even if students had learned the material they had not been encouraged to express what they had learned [because] the post-tests were not graded. . . . Few students made any attempt to integrate the three questions and to evaluate the responsibilities of individual moral agents as complex people. . . . We did not provide the students with any explicit instruction in conducting systematic moral analysis. We instead expected students to intuit the process by examining a series of cases. [350-351]

Given these deficiencies, it should not have been a surprise that the pre-test/post-test model did not work. To me, the last reason given is the most telling. The students might have been able to do good moral analysis by the end of the course, but if they were not given any practice or instruction in how to do an explicit, written moral analysis, it is not surprising that they failed to rise to the occasion when they were asked to do so at the end of the course. I can imagine that the students may have been able to do the analysis, but were simply unprepared to express it in the format required. This looks like a bad fit between course content and evaluation design.

After this initial failure, the instructors tried a different method for assessing student learning. Instead of replicating the pre-test exactly, they asked their students at the end of the course to do a meta-analysis of their pre-test.

- (1) We provided [three] short vignettes (Appendix C). Two contained issues of ethical importance (from the faculty’s perspective) and one did not.
- (2) More explicit instructions were given for the pre-test and students were told that a ‘high quality’ response to the pre-test/post-test was necessary to receive a ‘Pass’ in the class.
- (3) At the time of the post-test, students received back their pre-tests, with instructions to analyze how well they had responded to the pre-test (Appendix D). [351-352]

Unfortunately, the authors do not include item 2 (the more explicit instructions for the pre-test) with this article. They do provide the instructions for the post-test:

The purpose of this final exam is to help assess what difference this class has made in the way that you think, dealing with ethical problems.

The diagnostic test that you completed at the beginning of the term is attached.

- 1) Please review the case, the instructions you received at the beginning of term and your responses.
- 2) Analyze your initial response. Describe how your thinking has changed. Be sure to discuss

understandings or information that you have now that you didn't have at the beginning of the term.

This is your opportunity to consider how your thinking has changed. Please notice changes in HOW you think as well as any changes in WHAT you think. It may be that you reach the same conclusion now that you did in the beginning of the term, but that you think about the situation in a different way.

- 3) Please attach your diagnostic test to the final exam.

Please keep in mind that you are NOT being asked to repeat the assignment from the beginning of the term. You are being asked to analyze how you initially responded to that assignment. [362, capitalization in original]

The authors describe the new pre/post-test method as a success, but say, "As we did not involve external evaluators, the report of the results are anecdotal but compelling."

## The Defining Issues Test

Another tool widely used for assessing the achievement of one possible proximate goal specific to the responsible conduct of research is the Defining Issues Test (DIT), developed by the late James Rest and his colleagues. The DIT is a multiple-choice, standardized test which measures moral reasoning ability. It is intended to emulate the lengthy open-ended interview Lawrence Kohlberg used to develop his typology of levels of moral development, and there are many studies to indicate that the DIT succeeds in this quite well.

The DIT has obvious weaknesses.

1. It measures only moral reasoning. While improving moral reasoning skills is an admirable goal, it isn't necessarily the only goal or even the most important one for a curriculum in the responsible conduct of science.
2. To be useful as a measure of the outcome of a course, it would have to be used as a pre-test and post-test, which may not be effective if your curriculum has a short time-span.
3. If you do use the DIT, you have to be sure you don't pin too much on its results. Since it only captures moral reasoning, you would want to make sure that you use other measures for your other goals.
4. The DIT is only an instrument; it is not a curriculum. It can be used for assessment, but not really for teaching.

That said, the DIT has obvious strengths. It is objective, easily administered, and there has been a great deal of study on its use and effectiveness. It doesn't do everything (what single measure does?), but it does what it was designed to do extremely well.<sup>3</sup>

## Moral Reasoning in Scientific Research

I helped develop another tool with the help of one of James Rest's colleagues, Muriel J. Bebeau. Moral Reasoning in Scientific Research: Cases and Materials is an 80-page booklet

---

<sup>3</sup> A recent description of the DIT and its theoretical foundation can be found in Rest et al. 2000. For information on using the DIT: Center for the Study of Ethical Development, University of Minnesota, 206 Burton Hall, 178 Pillsbury Dr SE, Minneapolis MN 55455.

designed both to teach moral reasoning and to assess improvement in moral reasoning. It includes six case studies as well as extensive information on how to use the cases and a discussion of the theoretical underpinnings of the approach.

The same weaknesses apply as for the DIT, except the fourth. The booklet is not an entire curriculum, but it is designed to be used both in teaching and assessment. The booklet also has many of the strengths of the DIT, except that it is only quasi-objective and it has not been tested nearly so thoroughly – though, as the introduction explains, the approach upon which the booklet was modeled has been extensively studied.

Information on Moral Reasoning, and the booklet itself in PDF format, can be found on the Poynter Center's World Wide Web site at <http://poynter.indiana.edu/mr-main.shtml>.

## Conclusion

To the best of my knowledge (which is, no doubt, incomplete), the Moral Reasoning booklet and the method described by Elliott and Stern are the only widely disseminated materials designed specifically for assessing student learning in the responsible conduct of research. If you know of others (whether published or not), I would be delighted to hear about them.

Assessing student learning is not easy, but it is an essential aspect of responsible teaching. And assessing student learning in the responsible conduct of research is not impossible. There are tools available, and imaginative teachers can devise more tools of their own.

## Works Cited

- Bebeau, Muriel J., et al. 1995. Moral Reasoning in Scientific Research: Cases for Teaching and Assessment. Bloomington, IN: Poynter Center. <http://poynter.indiana.edu/mr-main.shtml>.
- Elliott, Deni, and Judy E. Stern. 1996. "Evaluating Teaching and Students' Learning of Academic Research Ethics." Science and Engineering Ethics 2:345-366.
- Rest, James R., Darcia Narvaez, Stephen J. Thoma, and Muriel J. Bebeau. 2000. "A Neo-Kohlbergian Approach to Morality Research." Journal of Moral Education 29(4):381-395.