Teaching Ethics in Biomedical Science: Effects on Moral Reasoning Skills

Elizabeth Heitman, University of Texas School of Public Health, Houston, Texas, USA

Patricia J. Salis, University of Texas School of Public Health, Houston, Texas, USA

Ruth Ellen Bulger, Uniformed Services University of the Health Sciences, Bethesda, Maryland, USA

Key words: Defining Issues Test, DIT, Ethics education, Evaluation, Responsible conduct of research

Academic institutions that train professionals play an important role in ensuring that trainees learn the ethical norms of their respective disciplines, and that they learn to behave ethically from the start of their professional lives. The National Institutes of Health requirement that funded research training programs include education in scientific integrity has made formal courses on the responsible conduct of research increasingly common in academic medical centers and research universities.

There is still no consensus on what constitutes the most appropriate subject matter, format, methods, or faculty for teaching the responsible conduct of research. The objectives of general courses on the responsible conduct of research and scientific integrity typically include increasing students' understanding of the norms of scientific practice, their recognition of ethically problematic situations in science, and their ability to analyze and respond to such situations in a morally mature manner. Courses vary in the specific content, the number of contact hours, the format (lecture, small-group discussion, video or web-based tutorials), and the instructors' professional background and ethical expertise. The effectiveness of available courses probably also varies. Studies of how students are affected by formal ethics courses in such disciplines as engineering, law, dentistry, medicine, nursing, journalism, accounting, veterinary medicine, and social work have found that course design influences the extent to which students' ethical reasoning skills change during the courses (1-3). Such evaluation in the area of scientific integrity, however, is still in its infancy.

The syllabi of courses on the responsible conduct of research in several institutions suggest that such courses present at least three different kinds of instruction to students. The first is the "how-to" of science, in which the practical, procedural dimensions of science, rather than its ethical dimensions, are the focus: how to devise an experiment, give a talk, or write a manuscript. The second kind of instruction relates to the rules, regulations, and professional norms articulated by the organizations in which scientists work, their professional societies, and/or the government: how to make experimental data available for use, how to address suspected research misconduct, and how to deal ethically with animal and human subjects. Ethical considerations are often addressed as an aspect of these practical issues. Lecture and individual reading assignments are effective mechanisms for teaching both of these traditional types of subject matter, and students' understanding and retention can be evaluated by an objective written (including computerized) or oral exam.

Corresponding author: Elizabeth Heitman, PhD, University of Mississippi Medical Center, 2500 North State Street, Jackson, MS 39216, 601-815-1134 (voice), 601-366-3262 (fax), eheitman@surgery.umsmed.edu.

The third type of instruction presented by these courses relates to students' ability to recognize the ethical aspects of problems that they encounter in their research, and their ability to address these issues in a considered way. This instruction involves their developing moral reasoning skills rather than simply comprehending information, and it frequently uses case discussion or problem-based learning. Two decades ago the Hastings Center Project on the Teaching of Ethics proposed three criteria for evaluating the effectiveness of such instruction: 1) whether the student understands the central concepts; 2) whether the student can make cogent oral and written ethical arguments; and 3) whether the student can recognize ethical problems and examine them rationally (4). This evaluation is typically conducted through a more subjective examination using actual case analysis, possibly in a written or oral exam, but ideally in a more interactive setting.

The Hastings Center Project emphasized that helping students develop skills to recognize and analyze ethical issues and stimulating their moral imagination are fundamental to the effective teaching of ethics. The Association of American Medical Colleges handbook, *Teaching the Responsible Conduct of Research through a Case Study Approach* (5), has also stressed the need to enhance students' ethical awareness and problem-solving skills in formal education on the responsible conduct of research. Ideally, the courses should have a positive effect on students' actual and future behavior, helping individuals avoid ethically problematic behavior and enhancing their ability to resolve unfamiliar ethical conflict appropriately.

After several years of teaching a formal course on the responsible conduct of research at the University of Texas Health Science Center at Houston, the course's organizers sought to assess its effects and to determine what outcomes could be evaluated formally. The course, The Ethical Dimensions of the Biomedical Sciences, originated in 1984 as an institutional response to an incident with a foreign graduate student that would have been considered plagiarism for a student schooled in the United States (6, 7). Consideration of the case highlighted the administration's and faculty's need to articulate the university's ethical expectations and to teach U.S. academic and professional standards to all students. The primary objectives of the course subsequently developed by Dr. Ruth Bulger, and later continued by Drs. Stanley Reiser and Elizabeth Heitman, have been to encourage students' interest in the ethical development and goals of science, and to teach students to prevent, recognize, analyze, and resolve ethical conflicts in the daily conduct of their work (8).

From the beginning, the course has used a combination of formal reading assignments, didactic lecture, and small-group case discussion to address a wide variety of issues in the responsible conduct of research. Its faculty have always included both ethicists and bench and clinical researchers from various disciplines, both as lecturers and as discussion leaders. Most are senior faculty. Since 1988, the course has been a requirement for graduation from the Graduate School of Biomedical Sciences, and it is an elective for graduate students in the School of Public Health. For the past four years, approximately 120 students have enrolled in the course each fall, including 90+ from the Graduate School of Biomedical Sciences' 22 degree programs, 10-15 students from the School of Public Health's 11 degree programs, and several post-doctoral fellows from the UT Medical School and MD Anderson Cancer Center. Students in biomedical sciences typically take the course in their first semester, while others often enroll in the second half of their formal graduate study.

Objective written examinations demonstrated that the course effectively enhanced students' knowledge and understanding of both the practical how-to of science and the rules, regulations, and professional norms of research that the course addressed. Written analysis in the final exam demonstrated students' ability to identify and consider ethical issues. Students' course evaluations also confirmed that most of them found the course valuable to their professional development. However, the faculty wanted to assess the more comprehensive effects of the course on students' professional attitudes and behaviors.

To affect students' current behavior and shape their future action, instructors of courses in the responsible conduct of research must have three things: 1) an effective way to teach desired behaviors; 2) an effective way to motivate students to adopt these behaviors; and 3) a reliable way to measure behavior change. In a broad literature review, we found no clearly identifiable, successful method for teaching ethical behavior or motivating students to act

ethically. While there has been work on how best to evaluate students' comprehension and retention of information related to ethical conduct, we found no generally accepted way to measure the presumed beneficial effect of ethics courses on behavior.

In the absence of accepted measures of behavior change and future practice, surrogate measures of the effectiveness of courses on the responsible conduct of research are needed. Bebeau (9) and her colleagues have developed a set of teaching materials for education in the responsible conduct of research that considers four psychological processes in the decision to act ethically: moral sensitivity (the ability to interpret a moral situation and the effects of various courses of action on the parties involved); moral reasoning (judgment about which course of action is right); moral commitment (intention to do what is right) and moral perseverance (the ability to follow through with ethical behavior). Their method of evaluating the effectiveness of courses that use the group's instructional materials assesses the essential components of students' moral discernment and moral reasoning.

Efforts to define, implement, and assess education in the responsible conduct of research in graduate science programs have parallels in medical education, where considerable work has been done on the teaching of professional ethics and the evaluation of such teaching. The effects of ethics courses on medical students' moral reasoning skills have been studied since the late 1970s (10). Such evaluations have linked different types of ethics education with changes in students' moral reasoning, and have suggested that case-based discussion can significantly increase students' moral reasoning ability.

The Defining Issues Test (DIT) is the instrument used most frequently to measure moral reasoning skills and the effects of education on moral reasoning. The DIT was developed by James Rest and colleagues at the University of Minnesota Center for the Study of Ethical Development (11). The test is a standardized, computer-scored test that is easily administered to groups. It is based on Kohlberg's theory of cognitive moral development, which considers the principle of justice as the highest moral good. The DIT presents six morally problematic scenarios; the subject ranks the importance of various moral criteria for judging how to act, then chooses a course of action. Scores are reported in terms of a P%, which measures the extent of "principled" reasoning behind the individual's assessment of the cases. Cross-cultural applications have found that DIT scores increase consistently with subjects' age and education level.

This study explored whether two offerings of our course on The Ethical Dimensions of the Biomedical Sciences had an effect on students' principled moral reasoning, as measured by the DIT.

Methods

Following an IRB-approved protocol, a total of 215 graduate students who were enrolled in The Ethical Dimensions of the Biomedical Sciences course were asked to complete the DIT at the beginning (before-course) and the end (after-course) of the 1997 and 1998 classes. Use of individual codes protected students' confidentiality. Computerized scoring by the University of Minnesota Center for the Study of Ethical Development generated P% scores.* The analyses used students' change scores — the after-course test score minus the before-course test score — as the data. A preliminary analysis of differences in change scores between the 1997 and 1998 classes (t-test, independent samples) was performed to determine whether it was possible to combine the data from the two classes. Next the effectiveness of the course in improving students' principled judgment by was tested directly analyzing whether their change scores differed significantly from zero (t-test, matched pairs). Finally, an analysis of variance (ANOVA) test was run to determine whether students' gender or country of undergraduate education (US or non-US) was related to differential change scores.

Results

One hundred seventy-two students (80% of the original 215 students) completed both a before-course and an after-course test, 95 students in 1997 (87% of 109) and 77 in 1998 (73% of 106) (Table 1). One or both tests from 14 of these 172 subjects were excluded from analysis based on scoring criteria used by the University of Minnesota Center for the Study of Ethical Development. The final sample therefore contained 158 students who had valid scores for both the before-course and the after-course tests. Change scores did not differ significantly between the 1997 and 1998 classes (t=-0.88,

p=0.38), so a combined analysis of the two classes was possible.

The primary analysis assessed the course's effect on principled judgment: It revealed that the students showed no significant after-course improvement in principled judgment, as measured by the DIT P% score (Figure 1, Table 2). Indeed, the pattern in six of the eight sub-groups (Figure 2) was for after-course scores to drop slightly.

Follow-up analyses of the influence on change scores of students' gender and location of undergraduate schooling indicated that neither gender nor location of education had a significant effect for the combined 1997 and 1998 courses (Table 3), for the 1997 students alone (Table 4), or for the 1998 students alone (Table 5). For the combined group and the 1997 group, there was no significant interaction between the gender factor and the location-of-schooling factor, but this interaction was significant in the 1998 group (Table 5). The 1998 data in Figure 2 suggest that this result arose from the distinctive pattern among men educated in the U.S. Their after-course scores declined somewhat, while those of both groups of women and of men not educated in the U.S. either improved very slightly or stayed essentially the same.

Conclusions

The finding that no significant change had occurred in P% scores after the course on the

	Combined 1997 & 1998 Classes	1997 Class	1998 Class
No. people who took at least 1 test	215	109	106
No. people who took 2 tests	172 (80% of 215)	95 (87% of 109)	77 (73% of 106)
No. test pairs sent for scoring	172	95	77
No. test pairs not used	14	11	3
Final no. people or test pairs	158 (92% of 172) (73% of 215)	84 (88% of 95) (77% of 109)	74 (96% of 77) (70% of 106)

Table 1. Composition of final study sample.
* one or both tests in pair purged by scorers for invalidity; one of pair purged by us due to absence of valid pair-mate; scorers failed to process test pair

Group	Change Score Mean (SD)	t Value *	p Value
1997 & 1998 Combined	-1.27 (11.75)	-1.36	0.17
1997	-2.05 (11.64)	-1.61	0.11
1998	-0.40 (11.89)	-0.29	0.78

Table 2. Statistical evaluation of course's effect on DIT P% scores: t-tests (matched pairs) of change scores (after-course minus before-course) in combined classes and in each class alone.

Source of Variance	Degrees of Freedom	F Value	p Value
Gender	1	0.21	0.65
Country of education	1	0.54	0.46
Gender X country interaction	1	0.90	0.34
Error	154		
Total	157		

Table 3. Statistical evaluation of effect of gender and location-of-schooling on DIT P% scores: analysis of variance of change scores (after-course minus before-course) in combined classes.

Source of Variance	Degrees of Freedom	F Value	p Value
Gender	1	0.16	0.69
Country of education	1	0.09	0.77
Gender X country interaction	1	0.25	0.62
Error	80		
Total	83		

Table 4. Statistical evaluation of effect of gender and location-of-schooling on DIT P% scores: Analysis of variance of change scores (after-course minus before-course) in 1997 class.

Source of Variance	Degrees of Freedom	F Value	p Value
Gender	1	1.75	0.19
County of education	1	0.97	0.33
Gender X country interaction	1	4.86	0.03
Error	70		
Total	73		

Table 5. Statistical evaluation of effect of gender and location-of-schooling on DIT P% scores: Analysis of variance of change scores (after-course minus before-course) in 1998 class.

responsible conduct of research was a surprising and frustrating outcome, given the course's perceived value within the university and the number of studies that report significant changes in students' moral reasoning skills after similar courses in professional ethics. Even more perplexing was that students in most sub-groups actually showed slight declines in P% scores after the course.

Upon reflection, the authors concluded that principled moral reasoning is only one of a number of skills and concepts that we hope to teach and foster in our course. Much of the material and related discussion in the course focuses on common conflicts and practical ethical strategies in research and collegial interaction. Rest and colleagues (12) noted in 1999 that Kohlberg's theories, and thus the DIT, address formal ethical structures of society, what they call macromorality, and do not illuminate the micromoral phenomena of personal, face-to-face interactions in everyday life. Thus these null findings suggest that it is essential to ask different questions or use different methods to evaluate the complex issue of the outcomes of the course.

The establishment and ultimate success of education in the responsible conduct of research will require effective means of assessing the impact of such programs on students' knowledge, awareness, and moral reasoning. Under the most recent proposal requiring such education in all Public Health Service-funded institutions, a wide variety of formats appear to satisfy the new credentialing standards. Suggested options range from semester-long academic courses to day-long workshops to hour-long web-based tutorials, to self-study reading programs. As academic research institutions develop the expertise needed to provide education in the responsible conduct of research, mechanisms must also be developed to assess the extent to which these different formats are effective in enhancing participants' moral reasoning skills. Recent observations reported by Bebeau and colleagues suggest that some apparently unchanged DIT scores may mask important differences in moral sensitivity and reasoning (13). Expanded use of the DIT should strive to uncover all significant changes in moral reasoning in order that academic courses can target their educational intervention appropriately.

However, if the objective of education in the

responsible conduct of research is to shape the behavior of researchers and to reform the culture of research, methods for evaluating such change must be developed, and instructors must learn how to present the rules, regulations, and professional norms of science in a way that motivates researchers to adhere to them.

Note

***** The Center generated the P% scores using its new system of validity checks, which should be considered when comparing these results to those of older studies.

Bibliography

1. Schlaefli A, Rest JR, Thoma SJ. Does moral education improve moral judgment? A meta-analysis of intervention studies using the Defining Issues Test. Review of Educational Research 1985; 55: 319-20.
2. Self DJ, Baldwin, DC Jr., Wolinsky FD. Evaluation of teaching medical ethics by an assessment of moral reasoning. Medical Education 1992; 26: 178-184.
3. Self DJ, Wolinsky FD, Baldwin, DC Jr. The effect of teaching medical ethics on medical students' moral reasoning. Academic Medicine 1989; 64: 755-759.
4. Callahan D, Bok S. Hastings Center project on the teaching of ethics. In: Callahan D, Bok S, eds., Ethics teaching in higher education. New York: Plenum Press, 1980.
5. Association of American Medical Colleges. Teaching the responsible conduct of research through a case study approach. Washington, DC: AAMC, 1994.
6. Bulger RE, Reiser SJ, Studying science in the context of ethics. Academic Medicine 1993; (Suppl. 3): 55-59.
7. Reiser SJ, Heitman E. Creating a course on ethics in the biological sciences. Academic Medicine 1993, 68: 876-879.
8. Bulger RE, Heitman E, Reiser, SJ. The ethical dimensions of the biological sciences. New York: Cambridge University Press, 1993.
9. Bebeau M. Moral reasoning in scientific research. Cases for teaching and assessment. Bloomington, IN: Poynter Center for the Study of Ethics and Assessment, 1995.
10. Self DJ, Baldwin DC Jr. Moral reasoning in medicine. In Moral development in the professions: Psychology and applied ethics. In: J Rest & D. Navares, eds., Hillsdale, NH: Lawrence Erlbaum Associates, 1994.
11. Rest JR. Development in judging moral issues. Minneapolis, MN: University of Minnesota Press, 1979.
12. Rest J, Narvaez D, Bebeau MJ, Thoma SJ. Postconventional moral thinking: A neo-Kohlbergian approach. Mahwah, NJ: Lawrence Earlbaum Associates, 1999: 15.
13. Bebeau MJ. Influencing the moral dimension of professional practice: Implications for teaching and assessing for research integrity (abstract). A Research Conference on Research Integrity, Bethesda. MD, November 19, 2000.

Figure 1. Mean DIT P% Scores Before and After C Combined 1997 and 1998 Classes

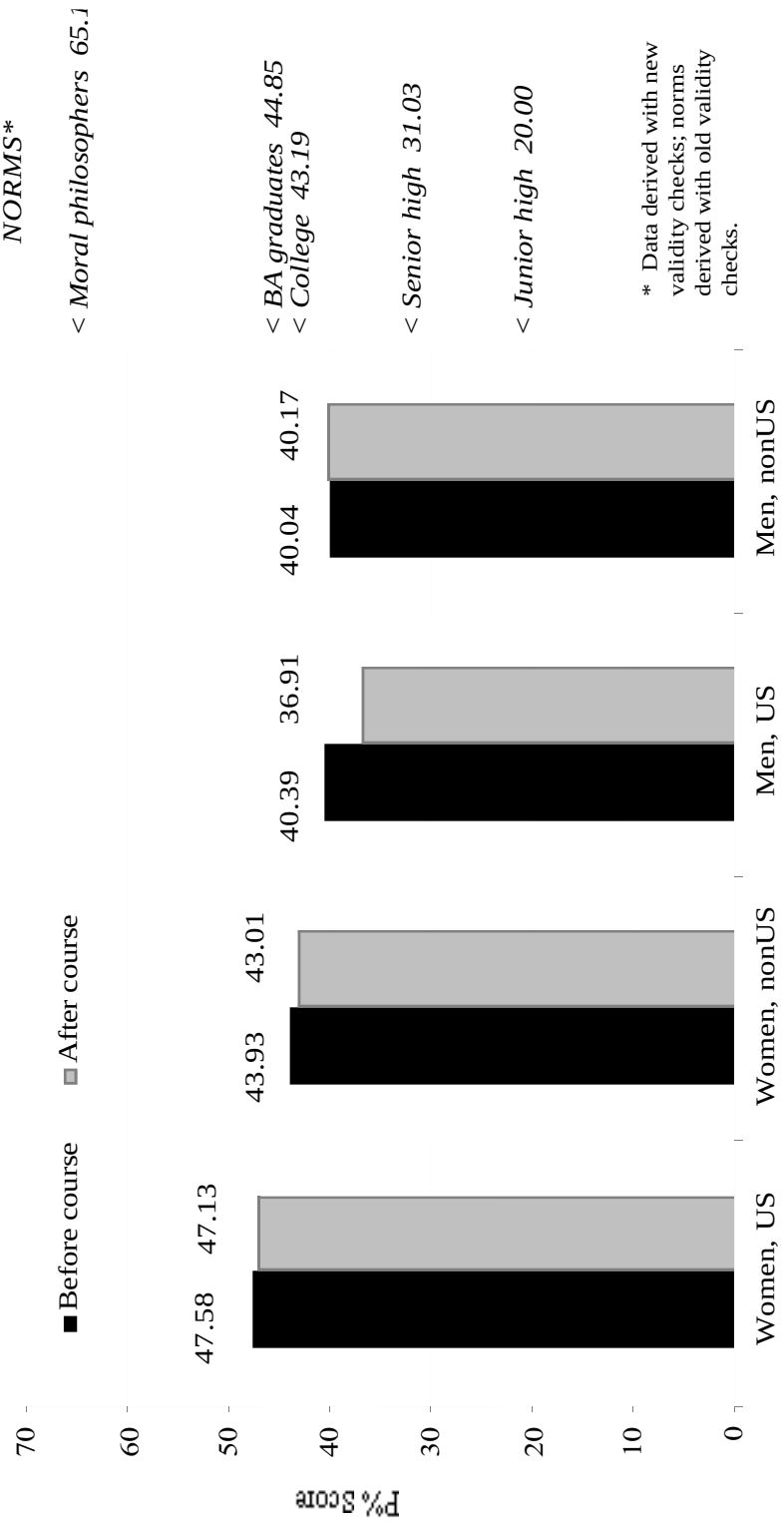

Figure 2. Mean DIT P% Scores Before and After C
1997 Class and 1998 Class