

## 5

# Assessment and Evaluation of Ethics Education and Mentoring

The following background questions provided a context for Session III, *Outreach and Assessment*:

Are relevant and important materials and techniques reaching the appropriate audiences? Who are the appropriate audiences, and are there useful feedback loops from them to the developers of materials, techniques, and guidance? Are the audiences able to adapt or adopt these resources? What efforts might improve access, use, and feedback and improvement? What kinds of assessment have been developed, make sense, or should be encouraged for the future? What have we learned, and what do we need to learn?

Felice Levine, executive director, American Educational Research Association, (AERA) moderated this session. Speakers were Melissa Anderson, professor, Department of Educational Policy and Administration, University of Minnesota Minneapolis; Daniel Denecke, head of the Best Practices and Publications Program, Council of Graduate Schools; and Joseph Whittaker, dean, School of Computer, Mathematical and Natural Sciences, Morgan State University.<sup>1</sup> The respondents were NAS member W. Carl Lineberger, professor, Department of Chemistry and Biochemistry, University of Colorado, Boulder; and Charles Huff, professor, Psychology Department, St. Olaf College.

---

<sup>1</sup> Brian Schrag, executive secretary, Association for Practical and Professional Ethics, had also been scheduled to make a presentation but was unable to attend.

One of the speakers in Session I, Michael Mumford, University of Oklahoma, also addressed the issue of assessment in reviewing the work of his research team, which compared results from its “sensemaking” training with other kinds of ethics training. Using a case-based pre/post measure, the team found that interactive “sensemaking” instruction had more positive results than some other approaches. Mumford reported that an evaluation of research-ethics courses at a number of research intensive universities showed that instruction given as part of regular classes that did not include interactive activities was generally not effective. In some cases, he said, this kind of instruction even had negative impacts on ethical decision making in four areas of research conduct—data management, the conduct of a study, professional practices, and business practices.

Melissa Anderson, University of Minnesota Minneapolis, reported on her research team’s survey of more than 7,000 early and mid-career NIH-funded scientists. Very few of the survey respondents reported that they had engaged in any fabrication, falsification, or plagiarism in the three years prior to taking the survey, but many indicated engaging in questionable research practices. A majority of mid-career scientists reported that they had cut corners or made inappropriate use of funds in those years. For both early- and mid-career scientists, the research indicated significant associations between these questionable practices and environmental factors, such as competitiveness, counter-norms (e.g., secrecy and self-interestedness), and perceived injustices in the research environment.

The survey results also indicated limited positive influence of ethics education on research behaviors, whether the instruction had been given in a separate course or was combined with other research training. The self-reports from early-career NIH-funded

scientists even indicated a negative relationship between separate ethics instruction and good data-handling practices.

In addition, the results indicated that the influence of mentoring depended on the type of mentoring. Mentoring focused on research ethics, good practice, and personal assistance was associated with a decrease in questionable behavior, but mentoring for survival (or how to get ahead in your field) was associated with an increase in questionable behavior.

Anderson recommended that laboratories and other research locations adopt a principle of “collective openness” that would require participants to encourage “anybody at any time [to] ask questions about any . . . work or how it is done . . . [and] raise questions so that mistakes, oversights, and misbehavior will . . . be caught..” Operating in accordance with this principle, she argued, would ensure that research behavior could “stand up to scrutiny” and meet the standards of “scientific integrity.”<sup>2</sup>

The next speaker, Daniel Denecke of CGS, reported that the 10 universities that participated in the first CGS project on ethics research (funded by NIH), found assessment to be a difficult challenge because of the difficulty of finding or developing measures of student learning. Denecke said assessments should also measure the institutional climate for integrity (which might explain differences between faculty and student perceptions) and the effectiveness of curricular reforms. The 10 participating universities assessed the effectiveness of efforts to get faculty buy-in rather than student learning.

The eight universities that participated in the second CGS project (funded by NSF), had some features in common, such as online modules, but they also developed their own

---

<sup>2</sup> Readers can find citations to this work at <http://cehd.umn.edu/EdPA/People/Anderson.html>.

activities and, especially, their own assessment strategies. Although a comparative assessment for these universities would have been helpful, Denecke said, the short lifespan of the project and the diversity of approaches had made that impossible.

He then described a new project that will have three layers of assessment. Measures of student learning will be left to the institutions, but the other two measures will be based on common instruments, one to assess student and faculty perceptions of cultural changes in their institutions and one to assess how well practices put in place for the project worked during the project and afterward, and to identify mid-course adjustments.

At various times in discussions throughout the meeting, workshop participants remarked that assessments of ethics instruction and mentoring were at an early stage of development, and that determining and adopting appropriate, consistent measures for success would not be easy. Even measures of student satisfaction and pre/post test achievement differentials, which are relatively easy to measure, do not tell if the right things are being measured or whether students can call on what they've learned afterwards, when needed. In addition, many assessment instruments have not been validated, and instructional methods may not always be appropriate for the target audience.

In the general discussion following this session, areas in need of further research, such as a multi-level assessment that would include individual outcomes and institutional changes over the short and long term, were identified. Among the commonly accepted, or at least usable, measures, the group named measures of broad-based faculty and departmental involvement at the institutional level, and measures of improvements in reasoning ability and other skills and knowledge at the individual level.

Some discussion participants noted that new, expanded, or revised programs offered by professional societies and accreditation bodies could provide another kind of measure. Felice Levine of American Educational Research Association suggested that questions might be embedded in ongoing periodic research surveys. For example, NSF could add an ethics question to its graduate student/postdoctoral survey. Several participants suggested that compliance officers in industry and academia might be asked to describe their experiences with different approaches to ethics education and to identify needs for further research.

The group was generally encouraged that attempts at assessment were being made and that the need for assessment has been recognized, if only in response to the new requirements of funding agencies, such as NSF. Many participants noted the urgent need for better assessment tools and a “menu” of choices to guide principal investigators who want to incorporate ethics training into their research programs, including assessments of training programs and “train-the-trainers” programs, to determine their consistency and effectiveness. Some members of one discussion group had floated the idea of national standards or certification but did not have time to pursue the idea in detail. Charles Huff of St. Olaf College also mentioned a variety of available measurement tools that might be adapted to ethics education, ranging from tests of personality, to those measuring recognition of ethical issues and knowledge of approaches to their resolution, to organizational ethical climate scales.<sup>3</sup>

---

<sup>3</sup> <http://www.stolaf.edu/people/huff/info/Papers/Good.Computing.P1.doc>