AC 2008-339: THE TEST OF ETHICAL SENSITIVITY IN SCIENCE AND ENGINEERING (TESSE): A DISCIPLINE-SPECIFIC ASSESSMENT TOOL FOR AWARENESS OF ETHICAL ISSUES

Jason Borenstein, Georgia Tech

Matthew Drake, Duquesne University

Robert Kirkman, Georgia Institute of Technology

Julie Swann, Georgia Tech

The Test of Ethical Sensitivity in Science and Engineering (TESSE): A Discipline-Specific Assessment Tool for Awareness of Ethical Issues

I. Introduction

There has been much written about the need for integrating ethics into the science and engineering curriculum. Efforts to accomplish this task are ongoing. However, assessing the effectiveness of ethics education programs generally, not just in science and engineering, has proven to be a rather daunting task.

Many of the attempts at assessment have made use of the Defining Issues Test (DIT), an instrument that measures moral reasoning based on Kohlberg's theory of moral development.[1] Briefly put, the DIT elicits subjects' responses to moral dilemmas and sorts those responses according to three types of moral reasoning: preconventional, conventional, and postconventional. A subject's responses are scored on the simple prevalence of postconventional reasoning, which involves reflecting on universal principles that apply to all of humanity, and also the prevalence of postconventional reasoning relative to the prevalence of preconventional reasoning, which corresponds to self-interest and the avoidance of punishment. Although there is scholarly debate about the merits of the DIT, among its advantages are that it is scalable and promises a quantitative measure of the effectiveness of ethics education.

Our own use of the second edition of the test (DIT-2) to compare different modes of ethics instruction at the Georgia Institute of Technology yielded troubling results: in a quasi-experimental study with pre- and post-tests and a control group, we found no statistically significant change in students' moral reasoning over a semester, even for those students who took a full course in engineering ethics.[2]

What the study did not tell us was whether this result was due to the ineffectiveness of ethics pedagogy or a shortcoming in the testing instrument. For a variety of reasons, we launched an investigation into the latter possibility. One option to consider is that the DIT-2 might be too general a measure of moral reasoning to capture the kinds of changes likely to be brought about by ethics instruction tailored to technical disciplines. Thus, we designed a new instrument for measuring moral reasoning that is patterned after the DIT-2, but with cases drawn from engineering and research contexts. We are still gathering data and analyzing the preliminary results from that instrument.[3]

At the same time, we considered the possibility that both the DIT-2 and our homegrown instrument were measuring the wrong thing. Following the research of those involved in developing the DIT-2, we note that moral judgment is only one component of ethical experience and conduct.[4] Along these lines, it is possible that instead of moral judgment, the primary benefit of ethics education may be that it enhances ethical sensitivity, the ability to identify and recognize relevant ethical issues emerging from a situation.

Consider, for example, the classic case in which a vendor offers a gift to an engineer who has authority in the hiring of vendors at a particular firm. To the practiced eye, accepting the gift would very likely bring the engineer into a potential conflict of interest. Yet students sometimes

have trouble seeing this as an ethical matter at all, casting as a matter of mere personal preference. In this light, ethics education can be seen as an effort to give students' eyes some practice, as it were. Case studies and other materials can serve to demonstrate, for example, how accepting what seems to be an innocent gift may lead to further temptations, which then lead in turn to a career-ending conflict of interest.

The idea of ethical sensitivity overlaps with some aspects of the idea of moral imagination, as developed by Mark Johnson, Patricia Werhane, and others.[5-7] One aspect of moral imagination concerns the ways in which people use conceptual schemata frame the situations in which they find themselves. For example, one person may frame a gift-giving situation as a potential conflict of interest while another frames it as a very nice perk of doing business. Moral imagination takes on a more critical function when it leads people to reframe the situation, either by taking up another person's point of view or by projecting narratives of what might happen next. In this light, ethics education can be seen as increasing students' stock of conceptual schemata and narrative possibilities, which would have the effect of making them more sensitive to the ethical dimensions of everyday situations.

Accordingly, one goal of ethics assessment could be to determine whether and what forms of ethics instruction would make students better attuned to the ethical dimensions of situations in which they may find themselves as students and as professionals.

Indeed, there have been a number of efforts to measure ethical sensitivity among students, particularly in professional degree programs.[8-12] The procedures used are often labor-intensive, however, calling for detailed, qualitative analysis of students' responses to situations that are presented to them. We set out instead to create a scalable instrument that may readily be used with large numbers of students with relatively little effort. The result is the Test for Ethical Sensitivity in Science and Engineering (TESSE).

II. Methods

A. The process behind the creation of TESSE

TESSE is a product of the collaborative efforts of three of the co-authors of this paper. The test consists of a set of seven case studies, each stated in a single paragraph. Once the cases and their associated lists of statements were generated, their arrangement was randomized in the test to avoid any sort of ordering bias. The cases are intended to reflect a situation that a scientist or engineer might reasonably expect to confront in professional practice. For example:

> Elena is part of a team working on a tight deadline to finish a public works project for SciEng Corporation, and the team had subdivided the tasks for efficiency. Elena's group has been developing theoretical models to predict the results of the system, while Matt's group has been working on a simulation of the system. When the two groups got together the day before the deadline, they found their two sets of predicted results were significantly different. Matt suggested that they go with the simulation results, because those outputs matched what they expected.

In the first version of the test, we included an open-ended production task after each case: subjects are asked to "comment on any or all professional ethical issues that you can identify in it." The test includes a large box in which the subject may respond in list, sentence, or paragraph form. In later versions of the test, our hope is that this open-ended task will fall away, leaving only the scanable and scalable survey instrument.

After each open-ended response, subjects are asked to complete a recognition task: the subject turns the page to find the case reprinted, but this time followed by a set of eight statements. The subject is asked to rate each statement on a five-point scale (from "Strongly Agree" to "Strongly Disagree") according to whether it "corresponds to an ethical issue in the [above] case." With the SciEng case, for example, goes the following statement: "Matt is being dishonest in not investigating the discrepancies".

The instructions for the test specify that subjects are not being asked whether they think the statements are true or false. Rather, subjects are instructed only to ascertain whether the statement reflects an ethical issue that appears within the case study in question.

In order to help ensure that what we are measuring is in fact ethical sensitivity, and to better detect degrees of refinement in the ethical sensitivity of each subject, we have introduced controls into the test at two levels. First, three of the seven case studies were deliberately created to be ethically neutral. The ability of subjects to detect this neutrality is indicated by their response to the following statement, which is presented after each case: "There is not a significant ethical issue in this case."

Second, each set of eight statements includes one or two statements that are intended to be important-sounding but are actually nonsense; they do not reflect an ethical issue at all, for example: "Jan has underestimated the potential for leveraging intellectual capital." This feature was designed in order to detect when subjects are trying to guess what the "right" answer is based only on linguistic cues.

B. Design of the study

We conducted the initial pilot study with TESSE during the Spring 2007 and Fall 2007 semesters, administering the test at the beginning and the end of the semester in five undergraduate courses and one graduate course.[*] Two of the courses (including the one graduate course) that did not include significant ethics-related content were used as the control group. Included in the experimental group were students enrolled in "Science, Technology & Human Values" (STHV)[†], "Science & Values in the Policy Process" (SVPP)[‡], "Ethics and the Technical

[*] Note that approval was obtained from Georgia Tech's Institutional Review Board (IRB) for human subjects research prior to the beginning of the study.

[†] The principal text for STHV was *Technology and the Future*, edited by Al Teich (10th edition).

[‡] The principal text for SVPP was *The Many Faces of Science: An Introduction to Scientists, Values, and Society* by Leslie Stevenson, Henry Byerly, and Leslie Stevenson. A collection of articles was also assigned for the course.

Professions" (ETP)[*], and "Moral Theories" (MT)[†] during the time when the study was being conducted. Faculty in the School of Public Policy at Georgia Tech instructed the courses in which students in the experimental group were enrolled.

The number of students taking the pre-test and post-test both in the control group and in the experimental group is displayed in Table 1. An identification number was assigned to students for the purposes of protecting their privacy. The number resulted from information about the course in which each student was enrolled along with randomized digits.

Course	Number of Respondents Completing Pre- and Post-Tests
STHV	28
SVPP	14
MT	30
ETP	195
Control group	131

Table 1: Participants who completed the TESSE pilot study by course

C. Scoring the TESSE

We have developed two indices thus far for scoring participants' responses on TESSE. The difference between the two measures centers on the relative weight given to various kinds of responses.

One index uses a likert-type scale, assigning points to each response. For statements that represent an ethical issue that is relevant to the given case, a score of +2 is assigned to a subject response of "strongly agree," +1 to a response of "weakly agree", 0 for neutral, -1 for "weakly disagree," and -2 for "strongly disagree". This scale is reversed for statements that do not represent a relevant ethical issue in the case. The respondent's final score is the average of the points that they have attained on all the issues for which they have provided a response.

The second index (denoted as *simple* in the analysis provided in the next section) follows the same logic, but differs in its point scale: +1 point for either "strongly agree" or "weakly agree" responses, 0 points for "neutral," and -1 point for "weakly disagree," and "strongly disagree" if the statement is a relevant ethical issue in the case. Again, the scale is reversed for statements that do not represent a relevant ethical issue.

III. Results

As of this writing, we are in the early stages of analyzing the data from our first runs of TESSE. We here present our current grasp of the results and their significance, but note that there is significant work to be done.

---

[*] The principal texts for ETP was *Engineering Ethics: Concepts and Cases* (3rd edition) by Charles E. Harris, Jr., Michael S. Pritchard, and Michael J. Rabins and *Creative Problem-Solving in Ethics* by Anthony Weston.

[†] The principal texts for the MT course were *The Elements of Moral Philosophy* by James Rachels (5th edition), and *The Right Thing to Do* by James Rachels (4th edition).

In our initial analysis of the data, examining the change in performance (post-test minus pre-test scores) between the experimental and control groups has not yielded significant results for the overall indices (see Table 2). The same can be said for the analysis of each individual experimental course group compared with the control group.

Overall Experimental vs. Control Group	N	Post - Pre Likert Score	Post - Pre Simple Score	Likert Difference p Value
Overall experimental group	267	-0.019 (0.016)	-0.014 (0.009)	0.264
Overall control group	131	-0.055 (0.028)	-0.016 (0.017)	Simple Difference p Value 0.906

Table 2: Comparison of post-pre scores on both scales between experimental and control groups on the TESSE. Number (N) is indicated, along with mean (standard error). [*, **, ***] denotes statistical significance at the [10%, 5%, 1%] level.

A scattering of significant results begins to emerge when we examine the data at a finer scale. For example, looking at the average scores on each of the seven cases yields some significant results, at least when we use the likert-type index.

Likert Differences by Cases	Likert Difference (Experimental - Control)	Likert Difference p Value
Case 1	0.061 (0.051)	0.253
Case 2	0.030 (0.052)	0.562
Case 3	-0.075 (0.078)	0.331
Case 4	0.069 (0.051)	0.177
Case 5	0.035 (0.049)	0.480
Case 6	0.144 (0.066)	0.040**
Case 7	0.000 (0.080)	0.998
4 Cases with Ethical Issues (1,2,4,5)	0.049 (0.026)	0.063*

Table 3: Comparison of post-pre likert scores between experimental and control groups for individual cases on the TESSE. The mean (standard error) indicated is the difference between the post-pre scores for the experimental and control groups. [*, **, ***] denotes statistical significance at the [10%, 5%, 1%] level.

In comparing the overall experimental group to the control group (see Table 3), for example, there is a positive effect (t-value) for case 6, significant at the 5% level. Also, when the four cases that include an ethical issue are considered together, they show a positive effect (t-value), also significant at the 10% level. Then again, case 7 yielded a result that was far from significant.

When we further disaggregate the results, examining each of the experimental course groups against the control, a pattern emerges as to which cases yield significant results.

For example, the ETP course deals with a wide range of issues in professional ethics for engineers and researchers, and TESSE yielded significant or near-significant results on more of the cases than with any other course group (see Table 4). The most significant result (at the 10% level) was on an ethically neutral case concerning the choice of materials (case 6). Two issue cases yielded results significant at the 20% level: one concerning data management (case 1) and one concerning mentorship (case 4). In the Spring 2007 semester, however, the instructor of the ETP course did not spend time on environmental issues, and nanotechnology was discussed only briefly. It might be expected, then, that analysis focusing on case 2, which concerned possible environmental risks from nanotechnology, did not yield significant results.

Likert Differences by Cases	Likert Difference (ETP - Control)	Likert Difference p Value
Overall Likert Score	0.040 (0.033)	0.235
Case 1	0.075 (0.053)	0.176
Case 2	0.021 (0.054)	0.698
Case 3	-0.065 (0.085)	0.441
Case 4	0.082 (0.054)	0.126
Case 5	0.038 (0.052)	0.473
Case 6	0.161 (0.071)	0.029**
Case 7	-0.021 (0.086)	0.804

Table 4: Comparison of post-pre likert scores between ETP course and control groups for individual cases on the TESSE. The typical sample size for ETP is 195 and 131 for the control group, but some students did not answer all cases. The mean (standard error) indicated is the difference between the post-pre scores for the experimental and control groups. [*, **, ***] denotes statistical significance at the [10%, 5%, 1%] level.

On the other hand, the version of the STHV course we studied in Spring 2007 did spend significant time on nanotechnology and on environmental issues, but not on a number of other

issues in engineering and research ethics. Comparing just that course group to the control group yielded near-significant results (approaching the 10% level) on case 2 (nanotech and environmental risk), but not on other individual cases. (Table 5)

Likert Differences by Cases	Likert Difference (STHV - Control)	Likert Difference p Value
Overall Likert Score	0.020 (0.064)	0.755
Case 1	0.120 (0.104)	0.156
Case 2	0.164 (0.102)	0.110
Case 3	-0.142 (0.140)	0.208
Case 4	0.074 (0.099)	0.453
Case 5	-0.040 (0.044)	0.695
Case 6	0.114 (0.134)	0.223
Case 7	-0.144 (0.152)	0.345

Table 5: Comparison of post-pre likert scores between STHV course and control groups for individual cases on the TESSE. The typical sample size for STHV is 28 and 131 for the control group, but some students did not answer all cases. The mean (standard error) indicated is the difference between the post-pre scores for the experimental and control groups. [*, **, ***] denotes statistical significance at the [10%, 5%, 1%] level.

It should be noted that all of these results tend to wash out when we use the simple index rather than the likert index.

Another finding of potential importance to our ongoing work with TESSE arises when we examine only the pre-test scores and group subjects by their response according to whether they reported having had previous ethics instruction. In the demographic data, we asked subjects whether they had had a "dedicated ethics course for the technical professions," a "general ethics or philosophy course", or "some ethics content in other courses."

We found significant difference at the 10% level in the pre-test scores on both indices (likert-based and simple) between those respondents who reported that they had taken a dedicated course on ethics in the technical professions and those who had not had such a course. The difference was significant at the 5% levels on both indices for students who had taken a full (dedicated or general) ethics course in the past and those who had no formal ethics course (table 6). We did not detect a significant difference between those who reported having some ethics content and those who reported no prior ethics experience.

Dedicated ETP Course vs. No Course	N	Pre Likert Score	Pre Simple Score	Likert Difference p Value
Dedicated ETP Course	90	0.842 (0.032)	0.485 (0.017)	0.098*
No Dedicated ETP Course	308	0.775 (0.020)	0.444 (0.011)	Simple Difference p Value: 0.064*

Full Ethics Course vs. No Course	N	Pre Likert Score	Pre Simple Score	Likert Difference p Value
Full Ethics Course	185	0.832 (0.023)	0.475 (0.013)	0.022**
No Full Ethics Course	213	0.754 (0.024)	0.435 (0.013)	Simple Difference p Value: 0.029**

Some Ethics Content vs. No Content	N	Pre Likert Score	Pre Simple Score	Likert Difference p Value
Some Prior Ethics Content	361	0.792 (0.018)	0.455 (0.010)	0.770
No Prior Ethics Content	37	0.775 (0.055)	0.443 (0.028)	Simple Difference p Value: 0.707

Table 6: Comparison of pre-test scores on both scales between students with various (reported) degrees of prior ethics experience on the TESSE. Number (N) is indicated, along with mean (standard error). [*, **, ***] denotes statistical significance at the [10%, 5%, 1%] level.

IV. Discussion

It is too soon to draw firm conclusions from these findings, but there are some promising signs.

Establishing the validity of a new assessment instrument is no trivial task. The apparent sensitivity of TESSE to prior ethics education and to the specific content of courses does seem to count in favor of the general validity of the approach we have taken, and perhaps in favor of the validity of at least some of the particular cases as we have designed them.

That said, there is clearly room for improvement. Case 7, a neutral case concerning the unproblematic response of an engineer to a directive from management, has yielded results that are frankly baffling. We are considering whether to drop it altogether. Further analysis of the data may give us grounds for revising or replacing other cases as well.

There is a further step we can take toward the validity of the test, which is to bring other experts from the fields of ethics, engineering, and research, into the revision process. While the three of us (two philosophers and an engineer) who developed the first version of TESSE were

able to come to an agreement, it would be useful to base subsequent versions on a broader and stronger consensus as to what is and what is not a significant ethical issue in a given case.

The task of developing a new instrument is further complicated in that we are devising the procedures for analyzing TESSE as we go. It seems clear enough that the simple index is not sensitive enough to be of much use, so we are for the moment focusing on the likert index. We plan also to devise an index that takes into account the variability of responses.

In the mean time, we continue to gather data, running the test with additional groups of students during the Spring 2008 semester. We welcome the participation of others who would like to try out TESSE in their own institutional contexts.[*]

References

1. Rest, J., et al., *DIT2: Devising and Testing a Revised Instrument of Moral Judgment.* Journal of Educational Psychology, 1999. **91**(4): p. 644-659.
2. Drake, M., et al., *Engineering Ethical Curricula: Assessment and Comparison of Two Approaches.* Journal of Engineering Education, 2005. **94**: p. 223-231.
3. Borenstein, J., et al., *The Engineering and Science Issues Test (ESIT): A Discipline-Specific Tool for Assessing Moral Judgment.* 2007, Georgia Institute of Technology.
4. Bebeau, M.J., *The Defining Issues Test and the Four Component Model: Contributions to professional education.* Journal of Moral Education, 2002. **31**(3): p. 271-295.
5. Johnson, M., *Moral Imagination: Implications of Cognitive Science for Ethics*. 1993, Chicago: University of Chicago Press. xiv, 287 p.
6. Werhane, P.H., *Moral imagination and management decision-making*. 1999, New York: Oxford University Press. x, 146 p.
7. Tivnan, E., *The Moral Imagination: Confronting the Ethical Issues of Our Day*. 1996, New York: Touchstone.
8. Bebeau, M.J., J.R. Rest, and C.M. Yamoor, *Measuring dental students' ethical sensitivity.* Journal of Dental Education, 1985. **49**: p. 225-235.
9. Brabeck, M.M., et al., *Increasing ethical sensitivity to racial and gender intolerance in schools: Development of the racial ethical sensitivity test.* Ethics & Behavior, 2000. **10**(2): p. 119-137.
10. Clarkeburn, H., *A test for ethical sensitivity in science.* Journal Of Moral Education, 2002. **31**(4): p. 439-453.
11. Hebert, P., et al., *Evaluating Ethical Sensitivity In Medical-Students - Using Vignettes As An Instrument.* Journal Of Medical Ethics, 1990. **16**(3): p. 141-145.
12. Hebert, P.C., E.M. Meslin, and E.V. Dunn, *Measuring The Ethical Sensitivity Of Medical-Students - A Study At The University-Of-Toronto.* Journal Of Medical Ethics, 1992. **18**(3): p. 142-147.

[*] Anyone interested in using TESSE should contact the Center for Ethics and Technology at Georgia Tech: www.ethics.gatech.edu.