

The Evaluation of a Computer Ethics Program

Graeme J. Byrne and Lorraine J. Staehr
La Trobe University, Bendigo, Australia

g.byrne@latrobe.edu.au l.staehr@latrobe.edu.au

Abstract

The importance of teaching computer ethics is shown by its inclusion as core content in both information systems and computer science model curricula. This paper outlines a method for evaluating undergraduate computer ethics programs using the Defining Issues Test of moral judgment. A “before-and-after with a control group” research design was used. The experimental group exhibited a significantly larger increase in moral judgment development than the control group. This indicates that computer ethics teaching is having a useful effect on students’ ethical development.

Keywords: computer ethics, Defining Issues Test, ethics education, moral development, professional ethics

Introduction

The importance of teaching computer ethics in undergraduate information technology (IT) degree programs is shown by the topic's inclusion as core content in the Computing Curricula 2001 body of knowledge (Engel, 2001), the Information Systems (IS) 2002 body of knowledge (Gorgone et al., 2002) and the Australian Computer Society (ACS) body of knowledge (Underwood, 1997). In Australia the teaching of computer ethics is mandatory for ACS professional level accreditation of an IT degree.

While it is important to develop effective teaching and learning strategies to obtain quality learning outcomes for students in computer ethics programs, it is equally important to design strategies to evaluate the success of these programs. The teaching and learning strategies that are used in the computer ethics component of an elective capstone course on social, ethical and legal issues in IT at La Trobe University, Bendigo are described fully in Staehr (2002). In this paper the evaluation of the computer ethics component of the course is reported.

Background

The Four Component Model of Moral Behavior (Rest, 1994) outlines the conditions necessary for an individual to behave morally (see Table 1) and provides useful guidelines for the design of moral education programs. The student must:

Material published as part of this journal, either on-line or in print, is copyrighted by Informing Science. Permission to make digital or paper copy of part or all of these works for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial advantage AND that copies 1) bear this notice in full and 2) give the full citation on the first page. It is permissible to abstract these works so long as credit is given. To copy in all other cases or to republish or to post on a server or to redistribute to lists requires specific permission from the publisher at Publisher@InformingScience.org

1. be able to recognize that a moral dilemma exists (moral sensitivity),
2. have the ability to make a morally justifiable decision,
3. place moral values above other values (professionalism), and

4. have the strength of character to carry out the morally justifiable course of action

for moral behavior to occur. This model is useful to consider when designing moral education programs so that all aspects of moral behavior are addressed. Therefore, although this study only evaluates moral judgment it is important that in computer ethics teaching the other three components of moral behavior are addressed. Bebeau (1994) and Duckett and Ryden (1994) provide ideas for learning activities and curriculum in professional education programs that address all four components in the model.

**Table 1: Four psychological components determining moral behavior
(adapted from Rest, 1994)**

Component	Explanation
Moral sensitivity (Interpreting the situation)	The awareness of how our actions affect other people and the requirement to decide between different courses of action.
Moral judgment (Judging which action is morally right/wrong)	What the DIT (based on Kohlberg's work, see explanation below) purports to assess—which action is more morally justifiable?
Moral motivation (Prioritizing moral values relative to other values)	What importance is given to moral values in relation to other values? Do values such as self-actualization or protecting one's organization replace concern for doing what's right?
Moral character (Having courage, persisting, overcoming distractions, implementing skills)	Psychological toughness and strong character is required to carry out a moral course of action. It is no good wilting under pressure.

This study used the Defining Issues Test (DIT) of moral judgment to evaluate the teaching of computer ethics to students enrolled in an elective capstone course on social, ethical and legal issues in IT. (Note that the questionnaire is copyright and may be purchased from the Center for the Study of Ethical Development, University of Minnesota). The theoretical basis of the questionnaire is Kohlberg's theory (Kohlberg, 1986) of moral development. The DIT has been used to assess moral education programs for a variety of professional/occupational groups, for example, medicine, dentistry, teaching, accountancy and journalism. The diversity of professions that have used the DIT indicates its likely successful application to the assessment of professional ethics programs in the information systems and computer science disciplines.

The DIT is not the only instrument that could have been chosen to evaluate computer ethics teaching, as it is not the only test that purports to measure moral judgment. An alternative contender is the Moral Judgment Test (MJT) also based on Kohlberg's work. It was developed in Germany and is widely used in Europe. There has been some discussion in the literature about the relative merits of each test (Lind, 2001; Rest, Thoma & Edwards, 1997).

The DIT was chosen rather than the MJT for three main reasons. Firstly, there has been more consistency in the version of the DIT studied over the years (i.e. a stable set of dilemmas, items and instructions to participants compared with the MJT). Secondly, the DIT has been used in a wider range of types of research studies. And thirdly, the original MJT is written in German and the equivalence of its translation to English has been questioned by Rest et al. (1997).

The standard DIT has generic moral judgment scenarios for students to score, not professional specific scenarios. Professional specific adaptations of the DIT have been developed, for example, for teaching, nursing, and journalism (Rest & Narvaez, 1994). An instrument modeled on the DIT with computer-related scenarios called the Ethical Dilemmas in Computing Test (EDICT) was developed by Bickel, Larrondo-Petrie & Bush (1992). Although the validation of an early version of EDICT against the DIT showed favorable results, the authors were unable to find evidence in the literature of a validated final version of EDICT. Hence they chose to use the standard DIT in the study.

The rest of this paper contains a method section that describes the subjects in the study, the research design, background theory and details about the DIT instrument, and the procedure used. The analysis and results are reported, followed by a discussion and conclusion section.

Method

Subjects

Of the 35 students who were the subjects in the study there were 20 in the experimental group. These were students enrolled in the course with a computer ethics component. The 15 students in the control group were final year IT student volunteers who were not enrolled in the course that had a computer ethics component. The demographics of the sample are shown in Table 2.

Table 2: Demographics of the Sample

Year	Gender	Experimental Group	Control Group
2000	Female	3	5
	Male	4	2
2002	Female	5	3
	Male	8	5
Total		20	15

Design

The experimental design used in this study was a repeated measure design where the within subjects factor was semester and the three fixed between subjects factors were year, group and gender. For further details on this type of design see Cobb (1998). The responses were the students' scores on the three indexes P, D and N2 (see the section on the DIT below for an explanation of these indexes). Ideally students should be randomly selected and then randomly allocated to the experimental and control groups. In this study, neither random selection nor random allocation was possible due to curricular restraints and the number of students available. The non-random selection may affect the external validity of this study while the non-random allocation may have implications for its internal validity. These issues are discussed in detail under the heading Experimental Validity and Model Assumptions in the Discussion and Conclusion section.

The DIT

The DIT was the instrument used in this study. It was developed more than 20 years ago at the University of Minnesota (Rest, 1979). The DIT is based on the work of Kohlberg (1986) who applied the cognitive developmental approach of Jean Piaget to the analysis of changes in moral

reasoning. Kohlberg presented his subjects with hypothetical moral dilemmas in an interview and analyzed their responses to assess moral development. He found from his interview data that moral growth begins early in life and proceeds in stages throughout adulthood. He defined six stages that comprise a developmental sequence that can be viewed in terms of six conceptions of how to organize cooperation (Rest, 1994). The six stages with a brief explanation of each stage are shown in Table 3.

Table 3: Six stages in the concept of cooperation (Rest, 1994)

Stage 1	The morality of obedience: Do what you're told.
Stage 2	The morality of instrumental egoism and simple exchange: Let's make a deal.
Stage 3	The morality of interpersonal concordance: Be considerate, nice, and kind: you'll make friends.
Stage 4	The morality of law and duty to the social order: Everyone in society is obligated to and protected by the law.
Stage 5	The morality of consensus-building procedures: You are obligated by the arrangements that are agreed to by due process procedures.
Stage 6	The morality of non-arbitrary social cooperation: Morality is defined by how rational and impartial people would ideally organize cooperation.

The DIT is a multiple choice test that is based on the general approach of Kohlberg. It was purchased from the Center for Ethical Development at the University of Minnesota for use in this study. The DIT is a paper-and-pencil test measure of moral judgment derived from Kohlberg's theory. Instead of scoring free responses to hypothetical moral dilemmas in an interview (as in the Kohlberg procedure), the DIT presents 12 issues after a hypothetical dilemma for a subject to rate and rank in terms of their importance. Hence the DIT data consists of ratings and rankings, instead of interview responses, that are then scored by computer. Instead of envisioning the scoring process as classifying responses into Kohlberg's six stages, the DIT analyses responses as activating three schemas. The scores represent the degree to which a subject uses the Personal Interest (Stages 2 and 3), Maintaining Norms (Stage 4), or Postconventional (Stages 5 and 6) Schema. The schemas have a close relation to Kohlberg's stages, yet they are different. As with Kohlberg's theory, the schema scores purport to measure developmental adequacy – in particular, how people conceptualize how it is possible to organize cooperation in a society. In short, the DIT is a measure of the development of concepts of social justice (Rest & Narvaez, 1998).

The DIT has been used in research for more than 20 years and its validity is well documented (Rest, Narvaez, Bebeau & Thoma, 1999). Criteria of relevance to this study are:

1. Differentiation of various age/education groups—studies show that 30% to 50% of the variance of DIT is attributable to level of education (Rest & Narvaez, 1998).
2. The DIT is sensitive to moral education interventions of more than 3 weeks. One review of over 50 intervention studies reports an effect size for dilemma discussion interventions to be 0.41 ('moderate' gains) whereas the effect size for comparison groups was only .09 ('little' gain) (Rest & Narvaez, 1998). The treatment effect size is defined as the difference between pre-test and post-test marginal means for the experimental group, minus the same difference for the control group, divided by the square root of the mean square error from the analysis of variance. The formula used for calculating this quantity is

$$E = \frac{(\bar{Z}_{EX,Post} - \bar{Z}_{EX,Pre}) - (\bar{Z}_{C,Post} - \bar{Z}_{C,Pre})}{\sqrt{MSE}} \quad (1)$$

In (1) above, E denotes the treatment effect size, \bar{Z} is a marginal mean, EX denotes experimental group, C denotes the control group, $Post$ denotes a post-semester mean, Pre denotes a pre-semester mean, and MSE is the mean square error from the analysis of variance used to assess the significance of effects. The treatment effect size is just a standardized two-factor interaction effect that allows comparisons to be made with established benchmarks.

3. The DIT is significantly linked to many 'prosocial' behaviors and to desired professional decision making. One review reports that 37 out of 47 correlations were statistically significant (Rest & Narvaez, 1998).
4. The DIT is equally valid for males and females (Rest & Narvaez, 1998).
5. The internal reliability of the DIT is shown in Table 4.

Table 4: Cronbach alphas of P and N2 indexes (Rest et al., 1999, p. 92)

	Cronbach alpha	
	P index	N2 index
1979 composite sample, n = 994	.76	.80
1995 composite sample, n = 932	.78	.83

6. The DIT contains some items that are written in a pretentious and complex way but are actually meaningless. An M (meaningless) score recommends rejection of questionnaires reported above the M cutoff ("Guide for the DIT", 1993).
7. The consistency check picks up subjects who are randomly marking circles or who do not discriminate items and mark items with the same rating ("Guide for the DIT", 1993).

There are three different scores that can be used to evaluate changes in moral judgment, the P index, the D index and the N2 index. The P index reflects changes in the higher stages 5 and 6 and the D index reflects changes in lower stages 2, 3 and 4. The N2 index reflects two types of information. The first is the way the subject ranks the post-conventional items (similar to P index) and the second is the ability of the subject to distinguish between the lowest stages and the highest stages. The second part of the N2 index arose from the observation that in intervention studies developmental advance is shown by acquisition of stage 5 and 6 items and also by discriminating stage 5 and 6 from stage 2 and 3 items (Rest et al., 1999, p. 96).

Procedure

The DIT was administered at the beginning (week 2 – pre-test) and end of the semester (week 13-post-test) to both the experimental and control groups with the aim of detecting any significant changes in moral judgment over the semester. As there was a computer ethics component within the course attended by the experimental group the hypothesis was that students enrolled in this course would exhibit a greater increase in score for moral judgment at the end of the semester compared with those in the control group.

A brief outline of the computer ethics component of the course follows. Further details including student feedback on the teaching and learning strategies can be found in Staehr (2002).

The computer ethics component was taught over 4 weeks of the 13 week semester. It included lecture and tutorial work on major ethical theories and discussion of ethical issues by students and exposure of students to the opinions of IT professionals, a lawyer and a member of the clergy on the same issues. Students were also required to complete an assignment where they applied ethical theory to resolve an IT related ethical dilemma from their personal experience (Kallman, 1992).

The students were asked to keep a reflective diary. The purpose of the reflective diary was to encourage reflection on the topics and issues covered in the course. The opportunity for reflection has been shown to be a characteristic of successful moral education interventions (Rest & Narvaez, 1994).

The students were introduced to Kohlberg's stages in moral development. In addition they were involved in class discussion of moral dilemmas (different from the dilemmas on the DIT) in terms of Kohlberg's theory. It has been shown that effect size is related to exposure to Kohlberg's theory (Schaeffli, Rest & Thoma, 1985).

Analysis and Results

Only 68 of the possible 70 questionnaires were returned. Four of the 68 questionnaires, 5.9%, were rejected on consistency checks. This outcome is within the bounds of 5% and 15% given by the Guide for the DIT (1993). This resulted in 64 questionnaires (shown in Table 5) being available for subsequent analysis.

A standard repeated measures analysis of variance was carried out for the three indexes with all main and relevant interaction terms included in the model. The three indexes were analysed for completeness, but the N2 index is the most important because it shows stronger trends on validity criteria and has slightly better Cronbach alpha internal reliability (Rest, Narvaez, Bebeau & Thoma, 1999). All p-values less than or equal to 0.05 were considered to indicate a significant effect. Since all factors in the model have just two levels, main effects are calculated as a first order difference between the means of each level of the factor concerned. For the group effect, the difference is relative to the controls; for gender, it is relative to males; and for semester, it is relative to the pre-semester measure. For the two-factor, interaction effects second order differences are used, and for the three-factor, interaction effect a third order difference is used. Only significant effects are discussed in detail since only they are indicative of genuine differences in the index values.

Main Effects

The semester main effect was only significant for the N2 index. Overall the students showed improvements in moral development across the semester with an increase of 3.96 ($p = 0.024$) on this index. The Year, Group and Gender main effects were not significant on any of the measures.

Table 5: Useable questionnaires

Gender	Experimental Group	Control Group
Female	16	14
Male	21	13
Total	37	27

Two-factor Interactions

The only significant two-factor interaction was the Semester \times Group interaction for the N2 index. The interaction plot in Figure 1 shows that students in the experimental group had an increase of

$$\bar{Z}_{EX,Post} - \bar{Z}_{EX,Pre} = 42.75 - 35.08 = 7.67$$

over the semester, while those in the control group had an increase of

$$\bar{Z}_{C,Post} - \bar{Z}_{C,Pre} = 36.48 - 36.22 = 0.26,$$

giving a significant effect of

$7.67 - 0.26 = 7.41$ ($p = 0.035$) for the N2 index. Using this absolute interaction effect for the N2 index and equation (1), we calculated the treatment effect size to be 0.82. This value indicates a strong gain in moral development (Rest & Narvaez, 1998).

There were no significant three-factor interactions on any of the measures.

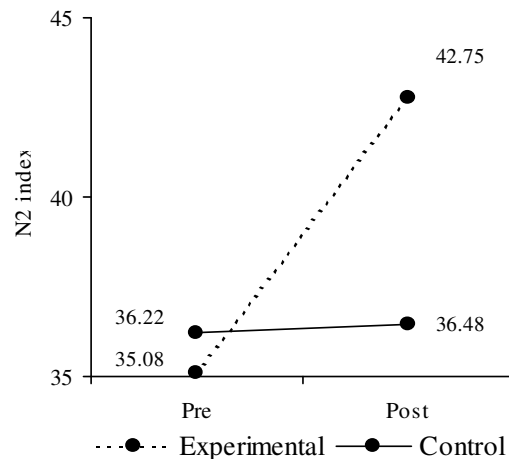


Figure 1: Group by Time interaction effect plot - N2 Index

Discussion

A review of 55 studies of educational interventions designed to stimulate development in moral judgment reported that treatments of 3-12 weeks, and those that involve dilemma discussion, all show larger treatment effect sizes (Schaepli et al., 1985). In this study the computer ethics component was four weeks, within the suggested treatment times. The computer ethics component involved dilemma discussion which is thought to improve students' moral judgment by exposing them to higher stage thinking from other students. The students were also exposed to Kohlberg's theory, another factor that has been shown to increase the success of moral education programs (Rest & Narvaez, 1994). The following quote from (Schaepli et al., 1985, p.343-344) discusses the effect of exposure to Kohlberg's theory:

“The impact of this exposure might be explained in two ways. One explanation is that reading the stage descriptions in effect instructs the subject how to perform on a test of moral judgment. A subject learns how to make a favorable impression by learning the theory. And so, exposure to the theory contaminates the post-testing. On the other hand, it might be argued that exposure to the theory is a powerful educational tool for actually changing a person's moral thinking. In this view, the theory facilitates restructuring of thinking, and the increase in posttest scores is not an artifact of an invalidated test, but a true indication of development. At this point, we cannot say which explanation is correct, and the issue deserves further research.”

However, there is evidence to suggest that exposure to Kohlberg's theory and dilemma discussion should not be interpreted as an attempt to "train" students to do well on the DIT. In a study by McGeorge (1975) different groups of subjects were asked to “fake good” (i.e. show the highest level of moral development) and “fake bad” (i.e. show a low level of moral development). The results showed that subjects asked to, were able to “fake bad”, but the group asked to “fake good” were unable to do so. This indicates that under normal conditions subjects are giving answers that reflect their actual stage of moral development rather than a higher “more desirable” stage.

The experimental group exhibited a significantly larger increase in the N2 index than the control group with a treatment effect size of 0.82 indicating a strong effect. The treatment effect size indicates that the computer ethics component resulted in an overall increase in moral judgment in the experimental group relative to the control group.

Experimental Validity and Model Assumptions

As mentioned earlier students were not randomly allocated to the experimental and control groups. Fortunately, this has not resulted in any problems with internal experimental validity as both the experimental and control groups had very similar average scores on the pre-measure. Also, standard residual analysis was carried out and no problems with the constant variance or normality assumptions detected. With respect to external validity, it must be remembered that random selection of subjects was not possible and the authors are therefore hesitant to extrapolate the results of this study to other student cohorts.

Conclusion

The results are encouraging and indicate a positive outcome due to the computer ethics teaching. However our results are based on small student numbers and restricted to one course in one university. To demonstrate a causal link between computer ethics teaching and moral development, further work is required. This further research could involve the evaluation of computer ethics programs using the DIT by our colleagues in other tertiary institutions.

References

- Bebeau, M. (1994). Influencing the moral dimensions of dental practice. In J. Rest & D. Narvaez (Eds.), *Moral development in the professions: Psychology and applied ethics* (pp. 121-146). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bickel, R., Larrondo-Petrie, M. & Bush D. (1992). EDICT for computer ethics education. *The Journal of Systems and Software*, 17 (1), 81-89.
- Cobb, G. W. (1998). *Design and analysis of experiments*. New York: Springer, 392-396.
- Duckett, L. & Ryden, M. (1994) Education for ethical nursing practice. In J. Rest & D. Narvaez (Eds.), *Moral development in the professions: Psychology and applied ethics* (pp. 51-69). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Engel, G. (Ed.) (2001). Computing curricula 2001 computer science. Retrieved December 18, 2003 from <http://www.computer.org/education/cc2001/final/cc2001.pdf>.
- Gorgone, J., Davis, G., Valacich, J., Topi, H. Feinstein, D. & Longenecker, H. (2002). Model curriculum and guidelines for undergraduate degree programs in information systems. Association for Information Systems. Accessed December 18, 2003 from http://www.spatial.maine.edu/SIEWEB/IS_ModelCurriculum.pdf.
- Guide for the Defining Issues Test. (1993) Version 1.3, Center for Ethical Development, University of Minnesota.
- Kallman, E.A. (1992). Teaching computer ethics through real life experiences. *Interface*, 14 (2), 56-59.
- Kohlberg, L. (1986). *Essays on moral development: The nature and validity of moral stages* (Volume 2). San Francisco: Harper & Row.
- Lind, G. (2001). Review and appraisal of the moral judgment test (MJT). Retrieved December 18, 2003 from http://www.uni-konstanz.de/ag-moral/pdf/Lind-2000_MJT-Review-and-Appraisal.pdf.

- McGeorge, C. (1975) Susceptibility to faking of the defining issues test of moral development. *Developmental Psychology*, 11 (1), 108.
- Rest, J. (1979). *Development in judging moral issues*. Minneapolis: University of Minnesota Press.
- Rest, J. (1994). Background: Theory and research. In J. Rest & D. Narvaez (Eds.), *Moral development in the professions: Psychology and applied ethics* (pp. 1-26). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rest, J. & Narvaez, D. (1994). Summary: What's possible? In J. Rest & D. Narvaez (Eds.), *Moral development in the professions: Psychology and applied ethics* (pp. 213-224). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rest, J. & Narvaez, D. (1998). *Supplement to guide for DIT-1*. Center for the Study of Ethical Development, University of Minnesota.
- Rest, J., Thoma, S. & Edwards, L. (1997). Designing and validating a measure of moral judgment: Stage preference and stage consistency approaches. *Journal of Educational Psychology*, 89 (1), 5-28.
- Rest, J. , Narvaez, D., Bebeau, M., & Thoma, S. (1999). *Postconventional moral thinking: A Neo-Kohlbergian approach*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Schaeffli, A., Rest, J. & Thoma, S. (1985). Does moral education improve moral judgment? A meta-analysis of intervention studies using the defining issues test. *Review of Educational Research*, 55 (3), 319-352.
- Staehr, L. (2002) Helping computing students to develop a personal ethical framework. *IEEE Technology and Society Magazine*, 21 (2), 13-20.
- Underwood, A. (1997). *The ACS core body of knowledge for information technology professionals*. The Australian Computer Society.

Biographies

Graeme Byrne is a Senior Lecturer in the Department of Mathematics at La Trobe University. He has taught in a wide variety of mathematical and statistical areas. His academic research interests revolve around approximation theory, statistical computing and analysis, experimental design, educational research, demographics and marketing research.

Lorraine Staehr is a Senior Lecturer in the Department of Information Technology at La Trobe University. Her research interests are in the business benefits of ERP systems, women in IT, and information systems education. She has published research papers in a number of international conferences and journals.