Assessing students ethical development in computing with the defining issues test challenges to be addressed

Suzy Jagger
School of Business and Social Science, Roehampton University
Email:s. jagger@roehampton.ac.uk

John Strain
Revd Dr John Strain, Lecturer in Ethics, University of Surrey, Guildford, GU2 7TE

This paper examines the early stages of an action research project designed to evaluate the pedagogic effectiveness of a teaching module in computing ethics. The project includes the use of the Defining Issues Test, an American designed paper and pencil test of students’ capability to make moral judgement. A comparison of the test results before and after the module suggested that some of the students had acquired new cognitive skills for ethical judgement and had rejected some of the more simpler approaches to ethical judgement they held hitherto. However the results also highlighted very low overall mean scores in what, in DIT terms, is called, post-conventional thinking skills. A wide variation between before and after scores was also apparent with some students seeming to benefit greatly and others regressing. The paper discusses these findings and, with the help of student interviews, identifies a number of factors which may have contributed to the results. These factors have provided rich insights into students’ ethical awareness and understanding. The results have contributed to the design of the action research project and have also enabled an appreciation of the strengths and weaknesses of assessing moral development in terms of cognitive factors.

Keywords: ethics, cognition, teaching, Defining-Issues-Test

INTRODUCTION

The aim of this paper is to examine some of the challenges to be addressed in measuring the success of teaching computing ethics. It draws on the early stages of an action research project to determine, firstly, to what extent teaching intervention succeeds in helping students develop ethical frameworks; and secondly to evaluate different teaching methods in their contribution to this goal.

The context of the project is the teaching of computing ethics in a Professional Issues module at Roehampton University, South West London, UK. This course is designed to help first year undergraduate computing students to understand the ethical and legal issues associated with the profession and to help them develop capabilities in moral decision-making to meet the types of ethical dilemmas they may encounter in future careers. The course exposes students to values and ethical concerns that they may encounter in future careers. The course exposes students to values and ethical concerns that are widely debated. But it is focused on scenarios that relate specifically to computing involving such issues as data protection and intellectual property. Teaching methods consist of group discussions of ethical scenarios designed specifically to be relevant to the student, step by step exercises in decision making, and classroom debates. Lectures focus on legislation as well as the application of philosophical perspectives such as deontology, utilitarianism and virtue ethics. Data to help determine the effectiveness of the various teaching methods include testing (the DIT), tutorial workshops, results from
assignments, and student interviews. It was considered that an implementation of the Defining Issues Test might provide some indication of student progress in appreciating, evaluating and responding to ethical dilemmas. Results have yielded some surprising findings and provided rich insight into student ethical awareness to aid in the design of the next phase of the project.

The article is divided into four sections. First, a brief review of previous attempts to measure ethical development is provided together with the significance of action research in understanding ethical development. Secondly, the particular challenges in assessing the acquisition of ethical capability in computing are addressed. Thirdly, some initial findings from an action research project are explained. And finally, the significance of the findings in appreciating the strengths and weaknesses of assessing moral development in terms of cognitive factors is described together with the lessons learnt for the future of the project.

**ASSESSING ETHICAL DEVELOPMENT AND THE SIGNIFICANCE OF ACTION RESEARCH**

Ideally, evaluating an educational contribution to ethical development would begin with finding a suitable method for measuring significant changes in students’ attitudes, affective responses and behaviours between the start and completion of the educational intervention. However, research has generated nothing that comes very close to measuring all of these changes together. In the field of medical ethics measurement is obtained through ‘observation of that behaviour over a period of time’ (QAA, 2002). For many educational courses the resources to measure changes and appropriate environments in which to observe students in practice do not exist. Considerable attention has been given to the relationship between moral behaviour and an individual’s level of moral sensitivity and judgement. Although moral reasoning skills alone are not sufficient to ensure high standards of conduct, they are, as Illingworth (2005) reminds us: “a necessary if not sufficient condition”. It is also widely accepted that these skills can be both taught and assessed. This confidence builds on the seminal work of Kohlberg and his theory of cognitive development in the moral domain, developed from his work in the 1950s with adolescent males. The theory identifies six stages representing three levels of thinking, pre-conventional, conventional and post-conventional. Very broadly, ‘pre-convention-
by Kantsians of dispositions, acquired through the education of our sentiments in moral practice. They differ in how principles and dispositions are related.

An important implication of this relationship between principles and dispositions is that the measurement of students’ development of principled ethical judgement will be significantly illuminated by attempts to understand the student’s perceptions of what their ethical development, in a Kohlbergian sense, actually means to them in their own life stories. For this reason, the evaluation of the learning module through the objective measure of the Defining Issues Test was set within an action research project in which the students’ own perceptions would figure.

**APPLYING THE DIT TO A COMPUTING ETHICS MODULE**

Discussions with computing undergraduates at Roehampton University indicate that they have above average levels of IT skills but often below average writing skills. The majority are male and from a wide variety of cultural and social backgrounds. The course contains a lower percentage of mature students than those on other courses and a low entry level compared with Computer Science courses. In discussions with students at the beginning of the course, they often complained to their teacher (and author, Jagger) that they didn’t understand why they should have to take this module in computing ethics (it is compulsory for all single honours computing students). Many do not think ethics is important or relevant for them, preferring to take another more computer-orientated module. This is the environment in which the course begins.

In such an adverse climate of opinion about the course the question arose as to whether significant learning could take place in the module. It was felt that a before and after model of study using the defining issues test might provide some useful information in relation to students’ moral reasoning capabilities and combined with a interview process afterwards it would help identify the students own perceptions of moral reasoning in their own moral lives. This might help elucidate the relationship between ethical reasoning and the acquisition of virtuous behaviour.

A primary method of testing, devised by Kohlberg himself, is the MJJ (Moral Judgement Interview) a tool in which a participant reads a selection of ethical dilemmas and is asked a series of questions, responses of which are categorised using the six stages. There are several variations on this dilemma based approach.

One such variation is the Defining Issues Test designed over twenty years ago by James Rest at the University of Minnesota. It utilises Kohlberg’s six developmental stages but, classifies them into three ‘schemas’: Personal Interest (PI), Maintaining Norms (MN) and Post-Conventional (P Score). Personal Interest reflects the extent to which an individual identifies an ethical choice with personal loss or gain. Maintaining Norms reflects the extent to which ethical choice is equated with conformance to what would be regarded as normal or conventional. Post conventional reflects the extent to which ethics is considered in relation to universalisable principles.

Rather than using free-responses to hypothetical moral dilemmas in an interview the DIT invites subjects to rate and rank twelve issues in terms of their importance in relation to a hypothetical dilemma. These rankings are then scored according to a scoring manual the results determining the degree to which a subject uses each of the three schemas. As with Kohlberg’s theory, the schema scores attempt to measure how people conceptualise social co-operation and it is therefore ‘a measure of the development of concepts of social justice’, (Bebeau & Thoma, 2003).

As well as the three schemas, Rest devised the N2 score to combine ‘the acquisition of new thinking’ (increase in P Score) with the ‘systematic rejection of simplistic thinking’ (decrease in PI score). The N2 attempts to measure the degree to which Post-conventional items are prioritised with the degree to which Personal Interest items are rated lower than Post-conventional items. The two parts are combined into one score by adding the P score to the rating data weighted by three. This score is used particularly for measuring the effectiveness of educational intervention.

The DIT has gained a lot of popularity over the years as a suitable measurement due to its stringent validity testing, as well as the ease with which it can be administered, making it particularly suitable for educational environments and for ‘before and after’ research models. Those particularly relevant to this study are briefly described here.

Boyd (1982) implemented the DIT on a professional business course and concluded that ethics based courses ‘significantly accelerate the rate of moral development’ however he was not able to identify which learning interventions were most effective in that study. Dukerich et al (1990) concluded moral reasoning scores are increased after group discussion – more so if the group leader has a high moral reasoning score. Cava (1990) reported...
an increase in DIT P-scores as a result of the introduction of discussion about deontological and utilitarian theories into the classroom. Such discussion emphasised the need for students to be able to examine and express their own views. She suggested critical thinking and group interaction are key components in ethics education.

Clarkeburn’s (2000) ‘before and after’ study amongst 195 Bioscience Students at the University of Glasgow revealed a lower P score and N2 score than the American average (in both the control and test group). Clarkeburn also revealed some regression in her P and N2 scores. She reasoned that the low mean and regression could have been a result of the changing nature of university cohorts and suggests that greater demands and pressures imposed on students now working to finance their degrees, and more goal-orientated teaching, means universities are not providing environments conducive for moral development to take place.

Wimalasiri (2001) also revealed lower P mean scores in a study on 106 Australian university students between the ages of 19 and 21. The results gave rise to questions as to whether the test is suitable for a non-American cohort due to cultural differences. Wimalasiri suggested that the dilemmas used in the DIT related somewhat exclusively to issues of right and legal justice. This ‘narrow domain’ precluded the exposure of the development of a broader definition of a moral field and impeded the revelation of progress to a higher level of reasoning. It follows that it may be incorrect to assume that subjects with lower scores, from different cultural environments, are ‘inferior’ in terms of the moral reasoning capability to their American counterparts. Having said this he also remarks that ‘the theory and the DIT appear to be an acceptable and reliable frame of reference for studying moral reasoning levels of individuals in different cultures.’ This indeed was the perspective of Kohlberg and Rest (Wimalasiri, 2001).

Rest describes morality as ‘a social construction, evolving from the community’s experiences, particular institutional arrangements, deliberations and the aspirations that are voiced at the time and which win the support of the community’ (1999:301). It follows that if this ‘common morality’ is a result of shared ideals and beliefs within a specific community, then the measurement of such a morality is better served locally using tools refined for that particular audience. Although he believed ethical theory should be robust enough to have universal application, to measure to any accuracy may require a more local approach. The DIT has over twenty years of documentation to support validity with submissions from over 10,000 respondents (Bebeau & Thoma, 2003:35) but the geographical location of submissions are primarily within the USA and Bebeau and Thoma acknowledge that even within this national spread, ‘caution should be used in interpreting these scores as data were not selected to counterbalance the variability that appears to exist among college and professional school programmes’ (2003:34). Although the scenarios chosen reflect universal issues (poverty, privacy, freedom of speech, euthanasia and repression), American primary and secondary school education is known to incorporate more overt representation of these issues than international counterparts.

Although there is no conclusive evidence to link moral development with improved ethical behaviour, there is evidence to suggest that learning does occur, at least in the short term (Wright, 1995) and that the moral development measurement techniques, although far from perfect, go some way into providing indications for educators on best practice.

**SOME INITIAL FINDINGS**

The Defining Issues Test was distributed to the students at the start of the twelve week module and was repeated at the end of the module. Forty four tests were completed and collected at the start of the module, six were removed due to incomplete answers and two more purged after analysis leaving thirty six ‘before’ students. Twenty nine tests were collected at the end of the module of which three were purged. Of this set of data, seventeen pairs of ‘before’ and after sets were able to be matched and analysed.

There were two main issues that emerged as a result of the implementation. The first related to the postconventional mean scores which were considerably lower (mean=23) than other studies for this age group. Previous research collected from the authors of the study based on over two thousand responses of first year US college students showed a mean score of 32.32 (Bebeau & Thoma, 2003). Such a large difference in score could be highlighting the different educational experiences between American and UK cohorts. Thus, contrary to Clarkeburn’s conclusion that low scores may be a result of an emerging student environment at university unsuited to developing moral reasoning skills, the DIT itself may simply be failing to deliver an accurate reading of moral judgement levels due to the widely different cultural, social and educational backgrounds of the UK respondents. This
conclusion is borne out in interviews with two students who experienced repeated low scores (described below) and who commented on both the language and structure of the test as being difficult to understand and follow.

The other result was a significant increase in the N2 scores for the seventeen matched pairs. This was particularly encouraging as the index was designed to measure educational intervention. Anonymous course feedback documentation suggested an increase in the N2 score was consistent with student experience as a number of student evaluations made reference to the course improving ability to think and develop skills in decision making.

Although the N2 scores showed a significant upward trend, it was nonetheless disconcerting to note inconsistent variations across the 17 subjects with 6 resulting in lower post-conventional scores than at the start of the course. In an effort to understand more accurately what may be happening with, three students were interviewed, one who exhibited a low score but showed significant improvement over the twelve weeks and two others who exhibited consistently low scores. The student who had achieved an increased p-score (10 to 20) reported that he enjoyed the course saying, “it helped me in gaining an understanding of things I need to think about not just at uni but afterwards when going into an IT career”. When questioned about why he thought he may have improved over the twelve weeks he stated, “Before I didn’t really have that much understanding so obviously as the course went through I learned quite a lot whereas if I’d had an understanding before the test wouldn’t have changed that much. Hadn’t really thought about it too much – professional issues in computing’ hadn’t really crossed my mind so I learned quite a lot from it.” He described the debates as “helping me learn the different viewpoints of people” and “put things in perspective”.

A second interviewee had a lower p-score after the twelve weeks, (P = 36 to 22). When questioned about whether the course had any impact on her dealing with ethical issues she replied, “it’s made me more politically aware of what different parties would suggest” and then cited two debates as being particularly helpful, the debate on children and computer games which highlighted her concern with regard to her little brother and the debate on illegal music downloading in which she stated, as a result of this debate, she would categorically not download music illegally considering the practice entirely unfair to the artists. The student made the following comments concerning the DIT;

“some of the words I didn’t actually understand – some of the political ones. I did find it quite a long thing and had to read through each one and wasn’t sure about that. Found it hard.” She described some of the wording as ‘peculiar’ and that “it didn’t sound as I would normally understand words in an exam.” She stated that she had other coursework hand-ins due in and, whizzed through it quickly as she’d already read it once twelve weeks earlier. When questioned about her experience in doing the ‘5 Step Method to Ethical Decision Making’ she stated, “I had trouble with that one. I had to think a lot and I found it really really difficult to look into myself and think what would I actually do in that situation? I found it very hard to put myself in the point of view of someone else.” These comments suggest that although finding the test difficult to understand, the student is also battling with the concepts associated with ‘principled thinking’.

In the case of the third student with a low score for both tests, (P = 14 to 8) when asked about the course she replied, “a lot of it went over my head to be honest…because it was so new and I didn’t have an interest in it at all so I did feel I was wasting my time. Some of it was helpful. You need to know your rights, copyright, different type of viruses, all that was good. In general it was a bit over my head but I did pick up some information that was useful”. When asked whether she understood the test she replied, “Yeah I did understand it and in some of the cases I put myself in that position and wow what would I do if I was in that position? If you were put in extreme cases what would you do? It did get you thinking.” But when asked about the second test she replied, “The first time round I done it properly but second time round just tried to remember what I’d done first time round. Some of it I didn’t really understand what they were saying and what they meant. It was a bit weird”. In both the last two cases, it should be noted that they were not told the results of the test until the end of the interview so as not to bias their comments. It was interesting to note that whilst student number 2 was very positive about the contents of the course and clearly got a lot out of it, student number 3 was not and only found some redeeming features upon reflection – yet both had a similar result from the DIT.

Clarkeburn (2000) revealed significant regression in her study and maintained that the high level of fluctuation in both directions on the DIT scale can be put down to students being profoundly confused about moral decision-making methods and thus, having had these methods introduced to them, employed the methods inconsistently. She states,
“Students were not only confused, but they predominantly used less sophisticated decision-making methods than expected for their age”. Clarkeburn (2000:34)

In fact the low scores could be highlighting students’ attempts to apply these approaches but failing to do them effectively due to lack of practice. This could suggest the need to build into the syllabus methods to facilitate further use of these skills in later years. It also suggests the need to ‘go slower’ in applying these methods, perhaps with more in-class exercises and formative assessment to ensure the concepts are implemented correctly.

CONCLUSIONS

This study has brought to light some important issues with regard to teaching and learning in computer ethics. First, student attitudes can be reasonably ambivalent about the importance of ethics. One possible implication of this at that when ethical development is measured using the DIT there is a risk that the results are being confounded to an uncertain extent by the students sense of the importance of ethics. It is not clear that Rist seriously considered this in the development of the DIT. There is no apparent innate acknowledgement of the importance of ethics in this field held by the students, regardless of the importance attached to it by the computing profession. It does not follow of course that the DIT cannot reflect significant changes in the development of ethical reasoning. But it is a reminder that there are a host of other factors, affective and motivational in the etiology of ethics.

Secondly, the overall post-conventional scores were extremely low for this student cohort. They were far lower than expected. The result supports the view expressed by some students that the sense of idealism inherent in post-conventional thinking is not shared by the students. This is consistent with a particular line of criticism of the Kohlbergian tradition. Not everyone is convinced, it would seem, that the individualistic idealism of post-conventional thinking is particularly desirable. Hogan (1973, 1975) argued that it is dangerous for people to place their own personalised principles above society and the law, however well meaning. The debate over the scientific merits of the DIT have been complicated by some ambivalence about the desirability of what the test assumes to be the most ethically desirable stage of development. More particularly, it may be important in the context of this research to understand what the attitudes of the students are to the very criteria upon which their ethical development is being measured.

Within the limited scope of the study, there is evidence suggested by the increase of N2 scores, that teaching ethics using group oriented approaches can have an impact on moral judgement capabilities. A key feature of the module was the incorporation of a debate as a learning exercise. The debate required everyone listen to different perspectives, to different possible answers to dilemmas, to different values with few constraints on what can be uttered. The DIT may provide useful indication of the pedagogic outcome of these debates in respect of ethical development. But it provided no indication of how the students views might have changed in relation to the key criteria of the DIT. To be specific, the students might have come to hold idealistic ‘post-conventional’ thinking in lesser regard, after the debate than before. What might be regarded as ethical ‘regression’ in DIT terms might in fact be educational progress in ethical reflection. At this stage, such considerations must be hypothetical. But they are pointers to the value of understanding not just progress in ethical principles but development in terms of students understanding of difference perspectives and their ability to empathise with others. Discussions with the students suggested teamwork such as debates, which requires students to take alternative viewpoints as well as working together for a common goal, inspire and motivate critical thought enabling moral decision making. It is for this reason that the action research project might profitably include a qualitative and ideographic appreciation of how the students regarded the debate.

The wording within the test itself is problematic for students from non US backgrounds, suggesting the DIT is less suited to a UK audience. Devising a UK equivalent test using the questionnaire/dilemma based approach, but within a more specific ‘UK’ (and indeed Computer Ethics) frame of reference may provide more accurate results and might be worth exploring further.

The next stage in the study will include a focus on the debates by collecting data after each debate relating to student understanding of ethical issues specifically with regard to computing scenarios to help differentiate the debate itself as a method for improving ethical understanding. This data can then be compared with DIT scores. This will help clarify the relationships between the universalisable world of Kohlbergian ethical reasoning and the incommensurable worlds of virtue ethics developed by individuals for particular practices and communities. It points towards a research agenda which is
both quantitative and qualitative. It also points towards an agenda which is rooted in the tradition of action research but one which combines quantitative and qualitative approaches. Knowledge is acquired from within the learning system rather than treating it as external the researcher’s world. The students are both subjects of tests and participants in qualitative research. The teacher is also a contributor to the quest for understanding of the significance of what is being taught in ethics.

REFERENCES


