

10-2004

# Cheating and Moral Judgment in the College Classroom: A Natural Experiment

Tim West

*University of Arkansas*

Sue Ravenscroft

*Iowa State University, sueraven@iastate.edu*

Charles B. Shrader

*Iowa State University, cshrader@iastate.edu*

Follow this and additional works at: [http://lib.dr.iastate.edu/management\\_pubs](http://lib.dr.iastate.edu/management_pubs)

 Part of the [Accounting Commons](#), [Educational Assessment, Evaluation, and Research Commons](#), and the [Science and Mathematics Education Commons](#)

The complete bibliographic information for this item can be found at [http://lib.dr.iastate.edu/management\\_pubs/12](http://lib.dr.iastate.edu/management_pubs/12). For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

---

This Article is brought to you for free and open access by the Management at Iowa State University Digital Repository. It has been accepted for inclusion in Management Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact [digirep@iastate.edu](mailto:digirep@iastate.edu).

---

# Cheating and Moral Judgment in the College Classroom: A Natural Experiment

## **Abstract**

The purpose of this paper is to present the results of a natural experiment involving academic cheating by university students. We explore the relationship of moral judgment (as measured using the defining issues test) to actual behavior, as well as the relationship between the honesty of students self-reports and the extent of cheating. We were able to determine the extent to which students actually cheated on the take-home portion of an accounting exam. The take-home problem was not assigned with the intent of inducing cheating among students. However, the high rate of observed cheating prompted the instructor to return to class and ask the students to provide information on their motivation. The students' responses are the data analyzed in this natural experiment. We found that in a simple regression the relationship between moral judgment scores and cheating behavior was insignificant. However, when we tested whether including Utilizer scores (i.e. the extent to which people select actions based on notions of justice) affected the relationship of cheating and moral judgment we found that Utilizer affected the relationship significantly. Finally, we found that moral judgment and honesty were not related, but higher levels of cheating behavior related to less honesty.

## **Keywords**

Accounting, accounting students, cheating, defining issues test, ethics, honesty, moral judgement

## **Disciplines**

Accounting | Educational Assessment, Evaluation, and Research | Science and Mathematics Education

## **Comments**

This is a manuscript of an article from *Journal of Business Ethics* 54 (2004): 173, doi: [10.1007/s10551-004-9463-x](https://doi.org/10.1007/s10551-004-9463-x). Posted with permission. The final publication is available at Springer via <http://dx.doi.org/10.1007/s10551-004-9463-x>.

Running Head: Cheating and Moral Judgment

## **Cheating and Moral Judgment in the College Classroom: A Natural Experiment**

March 17, 2004

Tim West  
University of Arkansas  
Department of Accounting  
Phone: 479 575 5227  
E-mail: [twest@walton.uark.edu](mailto:twest@walton.uark.edu)

Sue Pickard Ravenscroft \*  
Iowa State University  
396 Carver Hall  
Ames, IA 50011  
FAX: 515 294 3525  
Phone: 515 294 3574  
E-mail: [sueraven@iastate.edu](mailto:sueraven@iastate.edu)

Charles B. Shrader  
Iowa State University  
300 Carver Hall  
Ames, IA 50011  
Phone: 515 294 3050  
E-mail: [cshrader@iastate.edu](mailto:cshrader@iastate.edu)

\* Corresponding Author

# Cheating and Moral Judgment in the College Classroom: A Natural Experiment

## Abstract

The purpose of this paper is to present the results of a natural experiment involving academic cheating by university students. We explore the relationship of moral judgment (as measured using the Defining Issues Test) to actual behavior, as well as the relationship between the honesty of students self-reports and the extent of cheating. We were able to determine the extent to which students actually cheated on the take-home portion of an accounting exam. The take-home problem was not assigned with the intent of inducing cheating among students. However, the high rate of observed cheating prompted the instructor to return to class and ask the students to provide information on their motivation. The students' responses are the data analyzed in this natural experiment.

We found that in a simple regression the relationship between moral judgment scores and cheating behavior was insignificant. However, when we tested whether including Utilizer scores (i.e. the extent to which people select actions based on notions of justice) affected the relationship of cheating and moral judgment we found that Utilizer affected the relationship significantly. Finally, we found that moral judgment and honesty were not related, but higher levels of cheating behavior related to less honesty.

**Key Words:** accounting students, cheating, Defining Issues Test, ethics, honesty, moral judgment

# Cheating and Moral Judgment in the College Classroom: A Natural Experiment

## Introduction

Academic cheating is a type of behavior that is highly salient to educators but is problematic for experimentalists. Educators have immediate concerns about cheating because its occurrence distorts the assessment of learning. Furthermore, cheating may indicate that values considered essential to good citizenship and to good business practice have not been instilled. Cheating is almost universally condemned yet is widely engaged in, if self-reports are taken as credible (Moore, 1991). Whether or not cheating is on the increase is disputed by researchers (McCabe and Bowers, 1994; Spiller and Crown, 1995), but concern about the issue remains high (Crown and Spiller, 1998).

Some faculty members have become motivated to make the effort to teach ethics because of the relationship of behavior exhibited by college students and their behavior subsequent to graduate as they engage in business activities (Sims, 1993; Stevens and Stevens, 1987). This relationship has not been directly established, but some suggestive evidence exists. Researchers have found that college honor codes are associated with lower rates of cheating and that the existence of a college honor code has an impact on subsequent behavior in the workplace (McCabe et al., 1996). Mai-Dalton (1987) wrote that faculty cannot presume that college students will develop greater ethical sensitivity and awareness upon graduation without instruction and guidance. In the late 1980's a national commission studying fraudulent accounting reporting recommended more ethics training for students in order to help future business people analyze difficult situations more carefully (Johns and Strand 2000). Crown and Spiller (1998) argue that collegiate cheating could "fall within the rubric of unethical behaviors that are interrelated," (1988 684). Despite a lack of direct

empirical data tracking individuals and correlating their college cheating with subsequent unethical behavior on the job, Crown and Spiller observe that many theories of ethical decision making “treat unethical behaviors not only as interrelated, but also as unitary,” (1988 684). Thus faculty appear to believe that collegiate behavior is a precursor to or indication of subsequent professional behavior and therefore increasingly feel the need to include ethics in their curricula.

However, many faculty members introducing ethics face skepticism from colleagues and are plagued by the seemingly irresolvable question of whether ethics can be taught (Piper et al., 1993). Success in teaching ethics or any other topic is demonstrated ultimately through behavioral change. Thus the optimal form of assessment is observation and measurement of how people behave in situations where ethical choices must be made. However opportunities to observe such behavior are not under the control of researchers and, when they occur, tend to include a multitude of variables whose effect researchers cannot successfully disentangle. Creating behavioral responses in laboratories is another option; however, it is often difficult and sometimes morally unfeasible to test such behaviors in laboratory settings. Thus, determining an effective and convincing outcome assessment which relates directly to behavioral differences and which is feasible to administer becomes a crucial step in any effort to measure and improve ethics instruction. One such measure that has been adopted as an outcome assessment tool for ethical instruction is the Defining Issues Test (hence DIT), which is widely used to measure change in ethical judgment pre- and post- educational interventions.

The DIT is a paper and pencil test of moral judgment originally derived from a stage theory of moral development. The researchers at The Center for the Study of Ethical Development (hence Center) of the University of Minnesota who have developed the test

have changed their explanation of the theoretical basis for the test. They now discuss the test in terms somewhat different from those of stage theory. Instead, the researchers rely on schema theory and explain the scoring in terms of what it reflects about a respondent's reliance upon three schemas, the lowest being Personal Interest, the second being Maintaining Norms, and the highest level is called Postconventional. Previously the single score that was used most often in ethics research was called the P score and indicated what percentage of responses were in the highest (i.e. Post-conventional) stages. Recently, however the P score has been replaced by an N2 score which is more encompassing (Center, 1998; Rest et al. 1999).

Although the DIT has been tested extensively over decades and across many samples, the scope of the DIT is limited to moral judgment, one aspect of a four-component model that synthesizes much of the research on the psychology of morality (Rest et al., 1999, p. 100). The model's four components are moral sensitivity, moral judgment, moral motivation, and moral character. Because of the existence and wide-spread application of the DIT moral judgment is the only one regularly measured, although researchers are increasingly aware of the importance of the other three components in the model. The role of moral judgment, using the N2 Score from the DIT, in shaping ethical behavior has not yet been resolved. Thus, evidence that helps researchers understand the relationship of moral judgment, as measured on the DIT taken in the classroom, to actual behavior becomes more critical.

In this study data were obtained from a natural experiment which occurred when a faculty member gave students a take-home problem as part of an exam. Because of the ethical difficulty caused by any study involving inducement to cheat in the classroom, such behavior is rarely studied in its natural setting. Furthermore, a relatively small number of

laboratory studies have been conducted on cheating behavior. Reluctance to follow this particular line of research probably stems from a sense that such studies lack external validity because cheating in a laboratory carries fewer stigmas or has far less significance to a student than cheating on an exam in an actual classroom. Additionally, researchers may (and should) have some qualms about deliberately inducing dishonest behavior among subjects. Given the small number of studies involving cheating behavior among college students, there is a limited literature on the question of how moral judgment relates to the choices student make to cheat or not to cheat. Our study provides data on this important question.

## **Background**

The relationship of scores on the DIT to behavior is considered by Rest et al. (1999) to be an important criterion to assess the validity of the DIT; they cite over sixty studies showing a positive correlation between DIT scores and pro-social behavior in and out of the laboratory, in self-report, and in ratings by others (1999 81). Because, as we noted earlier, recent researchers often use the N2 score while earlier researchers used the P score, we are going to simply refer to “DIT score” to indicate the summary measure. If we are referring to a refinement or sub-set of the items on which the N2 is based we will be specific. Further attempts to validate the DIT, by determining its relationship to behavior and decisions, are ongoing. Not all studies show a positive relationship between moral or pro-social behavior and P scores (Thoma et al., 1991). Despite the somewhat inconsistent evidence for a strong and predictable relationship between moral judgment and behavior (Thoma et al., 1991), the DIT is used frequently to indicate that progress in developing ethical attitudes or learning has occurred in educational settings. The underlying assumption of such usage is that a changed DIT score signifies a change in behaviors beyond that of taking the DIT test more competently, i.e. obtaining a higher score on the DIT post-treatment.

## Literature Review

Academic cheating presents researchers with a very sensitive and rather delicate set of research ethics questions. Researchers are not allowed to entrap students into cheating and should probably be very cautious about even creating possibilities that encourage students to cheat. In cases where faculty members believe students have, in fact, cheated on actual exams or assignments, faculty are often reluctant to file complaints (Nonis and Swift, 1998). The infrequency of actual complaints filed by faculty belies the self-reports of students who say that cheating occurs regularly and frequently. Self-report of past cheating, of intentions to cheat, of reasons for cheating, and of attitudes toward cheating represent the dominant form of data on cheating (Spiller and Crown, 1995; Cizek, 1999). While such data are helpful and establish the seriousness of the cheating question, they do raise some concerns. There may be serious problems of accuracy in statements of intention; a wide gap may exist between what people say they do and what they actually do (Scheers and Dayton, 1987; Karlins et al., 1988; Miceli et al., 1991; Nowell and Laufer, 1997). There are issues of social desirability (and depending on the setting this can work to increase or decrease self-reported cheating); subjects may not believe that questionnaires are truly anonymous; or subjects may have little incentive to report honestly. In conclusion, actual behavior provides more validity than self-reports, but presents researchers with ethical challenges because of faculty members' reluctance to encourage cheating.

Lab studies offer an alternative to surveys and to observance of actual cheating, but are subject to limitations as well. Researchers must be cautious about inducing unethical behavior. Questions of generalizability may be more acute when socially undesirable behavior is under scrutiny (Crown and Spiller, 1998), as subjects may not want to appear to be unethical and the laboratory setting may not provide strong incentives for such behavior.

We are aware of two studies in the accounting ethics education literature relating ethical judgments to actual behavior. The first article addresses the extent of free-riding in the payment of course handouts. Ponemon (1993) describes an extensive effort to teach accounting ethics using a four-week module in one year and ten weeks of ethics instruction with the same students in the following year. Ponemon used two methods to measure the effectiveness of the instructional ethics intervention. He measured DIT scores prior to and after the ethics instruction, including the same two measures for a control group of students who did not receive the ethics instruction. He also tested the effectiveness of ethics instruction by asking students to contribute voluntarily and anonymously to the Accounting Department to pay for their printed course materials, as the department had used up its entire budget for handouts and would be unable to provide any additional materials unless students contributed. Although students believed their contributions would be anonymous, in fact they received bar-coded payment envelopes, which allowed the experimenter to determine how much each student contributed. Ponemon found that the ethics intervention, extensive though it was, had no effect on DIT scores. More interestingly, however, he found a parabola-shaped relationship of ethical behavior and DIT score. Free-riding (i.e. not contributing to the Accounting Department to pay the complete cost for materials already received) was highest when the DIT scores were in the lowest or highest quartile and lowest when DIT scores were in the middle quartiles. The average DIT score in the Ponemon study was 38, which is somewhat low for college students, and while Ponemon indicates that he broke DIT scores into quartiles, he does not give the ranges of scores in those quartiles.

Bay and Greenberg (2001) replicated certain facets of the Ponemon study. They asked for student volunteers who engaged in a trading experiment with conditions that allowed students to dissemble about the quality of what they were selling, and with economic

incentives (imposed sales quotas) which rewarded such dissembling. Bay and Greenberg did not replicate Ponemon's ethics instruction because their focus was entirely on the nature of the relationship of moral judgment and behavior rather than on the effect of ethics education. They found different behavior across gender. For the female subjects deceptive behavior increased as P scores increased; the higher a female's P score, the more likely she was to engage in deception to benefit herself. For males Bay and Greenberg found a quadratic relationship that replicated the pattern found in Ponemon; mid-level DIT scores corresponded to the lowest percentage of deception while higher and lower DIT scores corresponded to higher percentages of deception. Bay and Greenberg divided DIT scores into three ranges, with low indicating scores of less than 27, mid-range including scores between 27 and 41, and high indicating scores above 41 (2001, p. 374).

Like the Ponemon (1993) and Bay and Greenberg (2001) studies, our paper focuses on an activity that would, *prima facie*, relate to moral judgment, and is based on a justice orientation. Cheating is a violation of several norms, but primary among them is justice. The student who cheats has an advantage that other students do not have; thus the student who cheats can receive a grade that he does not deserve because he has created an unfair advantage. By cheating a student makes the playing field unlevel -- but not in acceptable ways such as studying conscientiously and preparing carefully -- but, rather, in unacceptable ways -- by gaining access to solutions through third parties or outside sources.

Looking beyond accounting, we find several studies of interest because they involve various measures of cheating. Newstead et al. (1996) say that no study has compared actual cheating at a university level to reported cheating. Newstead et al. argue that cheating behavior is driven primarily by morality and by achievement motivation (1996). However, they focus primarily on achievement motivation, do not use the DIT as an indicator of

moral judgment, and cite only two studies relating cheating to moral reasoning. Thus, the direct relationship of moral judgment to behavior or to the actor's description of his behavior has not been explored.

We located one paper involving actual cheating among business students which makes an indirect comparison of behavior to self-reports (Nowell and Laufer, 1997). These researchers report a two-stage study. In the first stage, they returned ostensibly ungraded quizzes to students to grade themselves. In fact, the instructors had recorded grades for the students but left no evidence of grading on the actual quizzes. They found that 23% of the students in five classes cheated at some time while grading their own quizzes. Then the authors used a subset of the same subjects to test the reliability of the Randomized Response approach to gathering survey data. The sample from the same students whose actual cheating rate was known to be 23% reported a much lower 9% rate, leading the authors to conclude that the randomized response results were not accurate. Because the second stage involved a sample of the population in the first stage, and because anonymity was guaranteed by the randomized response technique Nowell and Laufer could not directly test actual to reported cheating. Nowell and Laufer hypothesized that the reported rate of cheating was low because either the students did not truly believe their responses were anonymous or they did not believe that grading incorrect answers as correct constitutes cheating (1997).

In a study involving undergraduate students and their willingness to blow the whistle on research assistants who pressured the students to report incorrect data (Miceli, Dozier and Near, 1991) males were more likely than females to blow the whistle. More interestingly, the subjects' scores on the DIT were inversely related to their willingness to blow the whistle on the research assistants' unethical behavior.

Malinowski and Smith (1985) used a laboratory setting in which subjects were implicitly encouraged to exaggerate their performance by being presented with false, unrealistic norms and unfavorable comparisons to other subjects' scores. The experimenter absented himself during the experiment in order to allow subjects the opportunity to record the time they took to complete the task and the score they earned on the task. Of the 53 subjects 77% cheated at least once during the ten timed trials each completed. Malinowski and Smith found that subjects with higher DIT scores were less likely to cheat, or cheated fewer times, or took longer to begin cheating than those with lower DIT scores. Malinowski and Smith argue that the relationship of moral reasoning to behavior is "a matter of considerable theoretical and practical importance" (1985, p. 1024) and note that despite their fairly clear results showing that DIT scores related positively to pro-social behaviors, many aspects of that relationship are still unresolved.

Karlins et al. (1988) investigated cheating in a large sample of 1,374 students by determining whether or not they plagiarized on a single written assignment in a single course. They found a cheating rate of 3%, which is low compared to any other study of actual or self-reported behavior. They conjecture that possibly the results are unusual because the focus was on a single instance of cheating and measured only one possible type of cheating.

Gardner et al. (1988) used specially written study guides to determine whether students would, contrary to explicit instructions, rely on the study guide answers to complete their assigned homework, which accounted for 20% of the course grade in an introductory psychology course. They found that over a term approximately 50% of the students cheated at least once, though they found that students did not cheat consistently, i.e. to the same extent throughout the semester. They concluded that the determinants of cheating were "transitory rather than durational." (1988, p. 554). In none of the studies just cited do

students report back on how they behaved in a setting that would allow researchers to confirm the students' veracity.

## **Hypotheses**

Given that cheating is – at least in part – considered immoral because it violates fairness or justice in the classroom and because a greater number of previous studies find a positive relationship between pro-social actions or action-choices and DIT scores than find a negative relationship, we hypothesize that the relationship between DIT score and cheating is negative. In addition, given the fact that both studies involving accounting students (Ponemon, 1993; Bay and Greenberg, 2001) found quadratic (i.e. parabola shaped) relationships, we are going to test that type of relationship as an alternative to Hypothesis One.

**Hypothesis One:** The relationship of moral judgment (as measured by DIT) and cheating behavior is negative and linear.

Because researchers in the area of ethical behavior and judgment have found the relationship between moral judgment and behaviors to be less than straightforward, some intervening theoretical variables have been proposed. Thoma et al. (1991) suggest that when exploring the relationship between behavior and moral judgment, researchers should include the Utilizer (U) score, which indicates the extent to which people rely on justice reasoning as a basis for action decisions. Thoma et al. argue that the U score moderates the relationship of DIT score and action and the inclusion of the U scores in analysis helps to explain and rationalize previous mixed results. The U score is based on the action choices subjects selected in addition to the items students ranked as most important in making their choices.

The instructions distributed with the DIT state that including the U score has been shown to increase the explanatory value of moral judgment (Center, 1993; Rest et al., 1999, p. 104).

Marnburg (2001) suggests, however, that the relationship of DIT score to action is not straightforward. He argues that behavior results not simply from moral judgment but from a joint recognition of issues as moral and a willingness to act on those moral issues in certain ways. Marnburg questions whether people always act according to their best moral judgment and states that the relationship between specific types of reasoning and pro-social actions has not been clearly explained theoretically or convincingly demonstrated empirically. Marnburg tested the relationship of DIT scores to ethical attitudes in terms of policy and action choices in eight vignettes. He found that, contrary to his hypothesis, higher DIT scores were associated with greater inconsistencies between the subjects' policy choices and their individual action choices. He concludes that researchers must discard the notion of moral judgment or else define moral judgment in ways that provide more behavioral significance (2001, p. 282). However because of his doubts about the concept of the U score, Marnburg does not include it in his analyses. Thus, it is unclear if the inclusion of Utilizer would have altered the relationships Marnburg found. To offer some empirical evidence on these theoretical arguments regarding the usefulness of the Utilizer score as a moderating variable we incorporate the use of the Utilizer score in our second hypothesis.

**Hypothesis Two:** The relationship of moral judgments to moral behavior increases as the Utilizer score increases.

Finally, moral judgment may serve as a brake on unethical behavior by enabling people to recognize cheating as a violation of justice or fairness. We hypothesize that the same moral judgment that would inhibit cheating would also foster honesty. We test the

hypothesis that self-reports about cheating and actual cheating are more likely to coincide as DIT scores go up. In other words, we test whether honesty about one's actions is associated with higher levels of moral judgment.

**Hypothesis Three:** Honesty in self reports of behavior will be positively related to moral judgment.

## **Method**

While teaching an introductory managerial accounting course at a private university in the Midwest, one of the authors gave students a take-home problem as a portion of a mid-term exam. This University has an honor code which mandates that students refrain from cheating and that they disclose incidents of cheating which they observe. When the professor gave the assignment, students were told explicitly to not work with other students, although much of their earlier coursework had been done in teams. In addition, in class the students were told that they should not go to the Web or other computer sources to try to obtain help in completing the problem. In fact, the problem was selected from an Instructor's Manual provided by the publisher. However, the faculty member had removed the problem and its solution from the problems he placed on the Web site. In addition, the faculty member had explained the relevant technique for solving this type of problem to the students in a manner that differed significantly from the solution approach taken in the publisher-provided materials. Unbeknownst to the faculty member, another faculty member at the University had posted the intact Instructor's Manual, including the solution to the problem selected as the take-home portion of the examination, on the Web. The Web posting of the second faculty member included the textbook solution to the problem, albeit the solution differed significantly from what had been demonstrated to the students.

An interview with one of the students revealed that the assignment was given about halfway through the term when several other major assignments were due and midterms were being given. Thus, students felt rather stressed and tended to procrastinate on this accounting take-home portion of an exam. Some students decided, despite the professor's proscription, to work together and set times that they would meet. Other students chose to work alone but had friends who e-mailed the solution from the Web to them. Still others were aware that their classmates were not following the instructions to work alone and to avoid using Web sites.

When the students turned in the take-home portion of the exam, the faculty member quickly realized that solutions had been shared among students and that some students had obviously found the textbook solution on the Web as their answer matched the textbook and differed radically from the approach demonstrated in class. In assigning the problem the faculty member believed that very few students would be able to complete the entire problem, as it went beyond the material that had been covered in class, and planned to curve the grading accordingly.

The faculty member was initially very distressed to find a high level of cheating so a decision was made to address the issue directly by talking with the students and by asking them to respond to a series of questions to help the faculty member understand their motivation. Thus, students were asked to complete the DIT and to respond to a series of open-ended questions. The professor assured students that they would not face disciplinary charges and that the entire assignment would be eliminated when determining grades for the course. The students were assured that all responses were voluntary, although they were not confidential. The lack of confidentiality was necessary if the responses were to be matched to each individual student's actual behavior. However, the instructor assured the

students that complete anonymity in discussions or articles about the incident would be retained.

We determined the level of actual cheating in several ways. First, students were asked “To what extent did you receive assistance on the project from outside sources (e.g. working with friends, accessing the online answer manual)?” In addition, at this university a record is kept of all persons going to university Web sites, so the faculty member knew which students had gone to the other professor’s Web site to look at the Instructor’s Manual solution for the take-home problem. Finally, the faculty member could see which students had answers identical to those of other students in the class.

## **Subjects**

The subjects were 64 undergraduate students in an introductory managerial accounting course at a Midwestern university. Approximately two-thirds of the students were in the college of business, and nineteen were in the college of liberal arts. Age ranged very little as the undergraduate student population is almost entirely traditional. Because of incomplete responses by some students, the number of subjects whose responses are analyzed below was reduced to 55.

## **Results**

A contribution of this study is the use of actual, rather than experimentally-driven or self-reported behavior. We found that our measure was not as precise as would have been obtained in an experimentally controlled setting. To measure cheating we coded students’ responses to an open-ended question about help they obtained in doing the take-home question and categorized their responses into four levels of cheating. A 1 meant a student did not cheat at all; a 2 indicated that the respondent talked with another student; a 3 indicated that the student went to the Web site where the answer could be obtained; and 4

meant the respondent both talked with another student and went to the Web site. The coding was done by three people and any differences were discussed and resolved. The instructor also used the data obtained from the university regarding access to the Web site and compared solutions.

Students were categorized as cheating only if the evidence thereof was very clear. We believe this is important because there is research showing that students may have a rather narrow definition of cheating (Newstead et al., 1996), so student beliefs about cheating may differ considerably from faculty beliefs about what constitutes cheating. While we used a ranking that could imply that collaboration is less serious than obtaining solutions from the computer, we don't condone either of these approaches to take-home testing. However, our ranking does implicitly reflect students' comments about their behavior. They often mentioned that they were accustomed to and had been encouraged to work together and to help one another on homework and assignments. It strikes us as less likely that students could rationalize obtaining a solution from a Web site other than their instructor's Web site as the norm or as condoned behavior.

We found that some students claimed to have done less cheating than the professor claimed (under-reporters), while some student responses matched exactly to the professor's, and some students reported more cheating than the professor indicated. The latter situation could arise in several ways. Students might have looked on while a colleague logged onto a Web site. Only the person logging on would be considered a level 3 by the instructor, but the on-looker might have admitted to having accessed the Web. Students might have talked with other students and compared answers, but then not altered their responses or they may have intentionally made the answers look different. We chose to take a strict view of cheating by using the professor's evaluation when students under-reported and using the

student's evaluation when that evaluation matched the professor's or indicated greater involvement in cheating than the professor's assessment.

Overall we found that 26% of the students did not cheat at all. Forty-four per cent worked with a colleague and 30% obtained a solution from a web site, with two-thirds of those students sharing their information with another student. This is a rather disturbing result. In a situation where students were told that collaboration with other students and aids were not permissible, only one-quarter of the students complied. This was a situation where non-compliance was not a default action. Some students may have felt pressured to assist someone else; but students had to expend some effort to explore the internet and find the Web site that had the entire solution manual on it. In other words, this was not a situation where it was possible to cheat by accident or unintentionally. Given the nature of the situation, we felt that honest reporting by students was a significant indicator of their ethical development.

In our analyses involving honesty, we treated honesty as a binary variable. When students self-report matched or exceeded the category of cheating coded as described above, we considered the student to be reporting honestly. Given the strict standards we applied to categorize behavior as cheating, we considered those students who claimed to have done less than our coding indicated to be dishonest. Results of our statistical analysis follow.

**Hypothesis One:** The relationship of moral judgment (as measured by DIT) and cheating behavior is negative and linear.

We tested this relationship with a correlation of cheating (strictly defined as described above) and moral judgment, and a regression of DIT scores on cheating. The correlation was .175, which is not significant. The regression results were also insignificant (an F of .545 and a p value of .58). We also tested for a quadratic relationship between

cheating and DIT scores, as that type of relationship had been found in earlier work (Ponemon, 1993; Bay and Greenberg, 2001). The regression performed on a quadratic relationship resulted in a t-value of .592, which has a p value of .58.

**Hypothesis Two:** The relationship of moral judgments to moral behavior increases as the Utilizer score increases.

The U score can range from + 1 to - 1, with scores usually falling between .1 and .2. The mean of subjects in this study was .127, which is consistent with earlier studies. We tested the moderating effect of U by regressing the product of standardized DIT scores and cheating scores on U. The moderating effect of U on the relationship between moral judgment and behavior has been proposed by the developers and of the DIT test (Center, 1993, p. 24). The results of this test showed a standardized coefficient of -.252 with a t value of -1.823 and a p of .07. These results do not meet the .05 alpha test for significance.

**Hypothesis Three:** Honesty in self reports of behavior will be positively related to moral judgment.

When we analyzed honesty in relationship to DIT scores by using Fishers Protected Least Significant Difference test we found that there was no significant difference ( $p > .952$ ).

However, additional analyses resulted in findings of relationships that were of interest. We found the relationship of honesty in self reports to cheating behavior, using Fishers Protected Least Significant Difference test, to be significant and negative ( $p > .002$ ). Because lower coding for cheating indicates less cheating, a negative result implies that those students who cheated less were more forthcoming in their responses than students who cheated more. While it did not reach traditional significance levels, the correlation of

Utilizer score to cheating was .24, with a p value of .09, which is not in the direction we expected.

Although we did not hypothesize about gender effects, gender is often used in research on cheating (cf. Bebeau, 2002; Rest et al., 1999) and we performed some non-parametric analyses. We found that overall level of cheating was equivalent across male and female students, as was honesty in reporting after the fact; we should note that because of the small sample size power in these tests was limited. We include a table to demonstrate the distribution of cheating level and various scores on the DIT test scores sorted by gender and honesty in self-reporting.

**Insert Table One about here.**

Finally, Bebeau (2002) suggests that looking only at changes in the summary DIT scores may mask some patterns in other key variables that are now being provided by the Center for Study of Ethical Development. Following Bebeau's proposal that researchers refine their analyses by using scores on Personal Interest (which represent a pre-conventional stage of moral judgment) and Maintaining Interest (which represent the conventional stage of moral judgment) and the Postconventional schema (which is the P score and was used in earlier research as the primary measure of moral judgment), we substituted those scores in our first hypothesis and found no significant results. The p values for the simple regression of the schema score on cheating were .17 for Personal Interest, .91 for Maintaining Norms, and .23 for Postconventional.

When we substituted the individual schema scores into the analysis for the second hypothesis we found that U did not affect the correlation between Personal Interest and cheating (a standardized coefficient of .13 with a p value of .35) or the correlation between Maintaining Norms and cheating (standardized coefficient of .034 and a p of .81). However,

the correlation of Postconventional (P score) to cheating was significantly affected by the inclusion of the U score. The standardized coefficient was  $-.29$  with a  $p$  value of  $.04$ . This result implies that as Utilizer scores go up the correlation of P score and cheating declines or, conversely, that as Utilizer scores decrease the correlation of P score and cheating increases.

## **Conclusions**

Clearly any natural experiment is subject to limitations. Such studies result from circumstance and are perforce done without the extensive planning of a scheduled experiment. In our case the sample size is not large; only 55 students were involved and this episode relates to only a portion of a take-home exam. A different selection of students and differing circumstances may have resulted in different behavior. We found that only 26% of the students chose not to cheat by either collaborating with a colleague or accessing the solution on an unauthorized web site. Thirty percent of the students accessed the web site, and thereby contravened their instructor's guidelines for the take-home portion of their exam in an explicit and somewhat effortful manner. The remaining 44% worked with other students. While this behavior had also been proscribed, students often rationalized it by saying that they usually worked in groups and were accustomed to helping one another out in that way. While we find a rate of 74% to be disturbingly high, our results on the incidence of cheating are consistent with those of Malinowski and Smith (1985) reported earlier.

By contrast, our results relating the students' behavior to tests of moral judgment do not echo the data reported in Malinowski and Smith (1985). Our results showed that the DIT score on the Defining Issues Test had no significant relationship to cheating or to honesty about cheating. In our study, the responses to the DIT were provided only after

students had acted upon their decision regarding cheating. Therefore, we were not able to measure attitudes prior to the cheating incident. However, we believe these data provide an unusual and valuable view of actual, non-manipulated behavior among college students.

When we included Utilizer scores to test for a moderating effect, we found that higher U scores indicated a decreased correlation of P score (the Postconventional level of moral judgment) with cheating. This is surprising and contrary to what we expected based on Thoma et al.'s findings (1991) that inclusion of U strengthened the predictive value of DIT scores on behavior. To some extent our results duplicate those of Marnburg (2001) and provide rather worrying data. It appears that the subjects in both Marnburg's and our study who were most adept at answering the questions on the Defining Issues Test were also most adept at making action choices inconsistent with their stated policies or ethical choices. We believe these rather surprising results suggest that the relationship of U to the association of action and P score should be researched further.

We believe that the data from our natural experiment add to the research stream that shows little relationship of moral judgment, as measured by the DIT scores, to actual behavior in a setting that involves significant outcomes to subjects. We conclude that researchers must be cautious in drawing inferences from changes in DIT scores pre- and post-educational interventions; while the change in test score may be substantive and statistically significant, such changes may not correlate to changes in ethical behavior. The existence of a large body of mixed results may indicate a missing variables problem or a spurious correlation. Researchers need to explore what conditions are present when DIT scores and pro-social behaviors are correlated positively, and which conditions (internal or external to the subjects) are present when DIT scores and pro-social behaviors are not associated or are negatively related.

Researchers have shown that people tend to generalize negative impressions they have about others and to assume that if person X has been dishonest about one action, he or she is more likely to be dishonest in another way (Sanderson and Darley, 2002). Our results showing that students who cheated more were also more likely to be dishonest in self-reports about their cheating tend to support the validity of such generalization in this rather limited setting. Further research on classroom cheating should concomitantly examine the DIT with student motives, ethical climate, and other measures of the class environment. Only after examining cheating in its context can we hope to shed more light on these unexpected results relating to the DIT.

## References

- Bay, D. D. and R. R. Greenberg: 2001, 'The Relationship of the DIT and Behavior: A Replication', *Issues in Accounting Education* **16**(3), 367-380.
- Bebeau, M. J.: 2002, 'The Defining Issues Test and the Four Component Model: Contributions to Professional Education', *Journal of Moral Education* **31**(3), 271-295.
- Center for the Study of Ethical Development: 1993, *GUIDE for DIT-1* (University of Minnesota, Minneapolis, MN).
- Center for the Study of Ethical Development: 1998, *Supplement to the GUIDE for DIT-1* (University of Minnesota, Minneapolis, MN).
- Cizek, G. J.: 1999, *Cheating on Tests: How to Do It, Detect It, and Prevent It* (Lawrence Erlbaum Associates, Mahwah, NJ).
- Crown, D. F., and M. S. Spiller: 1998, 'Learning from the Literature on Collegiate Cheating: A Review of Empirical Research', *Journal of Business Ethics* **17**, 683-700.
- Gardner, W.M., J. T. Roper, C. G. Gonzalez, and R. G. Simpson: 1988, 'Analysis of Cheating on Academic Assignments', *The Psychological Record* **38**, 543-555.
- Johns, S. K., and C. A. Strand: 2000, 'Survey Results of the Ethical Beliefs of Business Students', *Journal of Education for Business* **76**, 315-320.

Karlins, M., C. Michaels, and S. Podlogar: 1988, 'An Empirical Investigation of Actual Cheating in a Large Sample of Undergraduates', *Research in Higher Education* **29**(4), 359-364.

Mai-Dalton, R.: 1987, 'The Experience of One Faculty Member in a Business Ethics Seminar: What Can We Take Back to the Classroom?', *Journal of Business Ethics* **6**, 509-511.

Malinowski, C. I., and C. P. Smith: 1985, 'Moral Reasoning and Moral Conduct: An Investigation Prompted by Kohlberg's Theory', *Journal of Personality and Social Psychology* **49**(4), 1016-1027.

Marnburg, E.: 2001, 'The Questionable Use of Moral Development Theory in Studies of Business Ethics: Discussion and Empirical Findings', *Journal of Business Ethics* **32**, 275-283.

McCabe, D. L., and W. J. Bowers: 1994, 'Academic Dishonesty among Males in college: A Thirty Year Perspective', *Journal of College Student Development* **35** (January), 5-10.

McCabe, D., L. Trevino, and K. Butterfield: 1996, 'The Influence of Collegiate and Corporate Codes of Conduct on Ethics-Related Behavior in the workplace', *Business Ethics Quarterly* **6**, 461-476.

Miceli, M. P., J. B. Dozier, and J. P. Near: 1991, 'Blowing the Whistle on Data Fudging: A Controlled Field Experiment', *Journal of Applied Social Psychology* **21**(4), 271-295.

Moore, M.:1991, *Cheating 101: The Benefits and Fundamentals of Earning the Easy "A."* (Moore Publishing, Hopewell, NJ).

Newstead, S. E., A. Franklyn-Stokes, and P. Armstead: 1996, 'Individual Differences in Student Cheating', *Journal of Educational Psychology* **88**(2), 229-241.

Nonis, S. A., and C. O. Swift: 1998, 'Deterring Cheating Behavior in the Marketing Classroom: An Analysis of the Effects of Demographics, Attitudes, and In-class Deterrent Strategies', *Journal of Marketing Education* **20**(3), 188-99.

Nowell, C., and D. Laufer: 1997, 'Undergraduate Student Cheating in the Fields of Business and Economics,' *Research in Economic Education* **18**(1), 3-12.

Piper, T., M. Gentile, and S. Parks: 1993, *Can Ethics Be Taught?* (Harvard Business School, Boston, MA).

Ponemon, L.: 1993, 'Can Ethics be Taught in Accounting?' *Journal of Accounting Education* **11**,185-209.

Rest, J., D. Narvaez, M.J. Bebeau, and S. J.Thoma: 1999, *Postconventional Moral Thinking: A Neo-Kohlbergian Approach*, (Lawrence Erlbaum Associates, Mahwah, NJ).

Sanderson, C. A. and J.M. Darley: 2002, "I am Moral, but You are Deterred": Differential Attributions about Why People Obey the Law,' *Journal of Applied Social Psychology* **32**, 375-405.

Scheers, N. J., and C. M. Dayton: 1987, 'Improved Estimation of Academic Cheating Behavior Using the Randomized Response Technique,' *Research in Higher Education* **26**, 61-69.

Sims, R. L.: 1993, 'The Relationship between Academic Dishonesty and Unethical Business Practices,' *Journal of Education for Business* **68**, 207-211.

Smith, K. J., J. A. Davy, D. L. Rosenberg, and G. T. Haight: 2002, 'A Structural Modeling Investigation of the Influence of Demographic and Attitudinal Factors and In-class Deterrents on Cheating Behavior among Accounting Majors,' *Journal of Accounting Education* **20**, 45-65.

Spiller, M. S., and D. F. Crown: 1995, 'Changes over Time in Academic Dishonesty at the Collegiate Level,' *Psychological Reports* **76**, 763-768.

Stevens, G. E. and F. W. Stevens: 1987, 'Ethical Inclinations of Tomorrow's Managers Revisited: How and Why Students Cheat,' *Journal of Education for Business* **63**, 24-29.

Thoma, S. J., J. R. Rest, and M. L. Davison: 1991, 'Describing and Testing a Moderator of the Moral Judgment and Action Relationship,' *Journal of Personality and Social Psychology* **61**, 659-669.

**Table One**  
**Descriptive Statistics**

	<b>Honest Male</b>	<b>Honest Female</b>	<b>Dishonest Male</b>	<b>Dishonest Female</b>
Count	19	19	11	6
Cheating Level <sup>1</sup>	2.5 (1.1)	2.4 (1.3)	3.5 (.5)	3.3 (.5)
N2 Score <sup>2</sup>	34.7 (11.2)	35.6 (12.8)	38.7 (12.6)	29.5 (4.3)
Utilizer Score <sup>3</sup>	.106 (.16)	.123 (.15)	.16 (.13)	.162 (.10)
Post- Conventional <sup>4</sup>	33.00 (11.1)	34.37 (12.3)	37.98 (12.9)	29.45 (3.9)
Maintain Norms <sup>5</sup>	32.1 (13.0)	29.2 (12.7)	31.2 (13.1)	35.6 (13.7)
Personal Interest <sup>6</sup>	33.00 (19.8)	27.5 (12.1)	25.9 (8.5)	28.1 (6.4)

Numeric Cells other than those in the “Count” row contain means and (standard deviations).

<sup>1</sup> A lower number indicates less cheating. The Cheating Level is 1 if no cheating occurred, 2 if the student talked with another student, 3 if the student went to the Web site, and 4 if the student talked with another student and went to the Web site.

<sup>2</sup> The N2 score has replaced the P score that was used previously (cf Rest et al. 1999 for a book length discussion of the reasons for the change).

<sup>3</sup> Utilizer scores can range from -1 to +1 but are generally in the .1 to .2 range.

<sup>4</sup> Post-conventional is the P score, which was formerly the most widely-used single measure from the DIT. A higher score represents a greater weighting on the highest stages of moral reasoning.

<sup>5</sup> A higher score for Maintaining Norms represents a greater weighting on the conventional (mid) stage of moral development.

<sup>6</sup> A higher score represents more responses in the pre-conventional or lowest stages of moral judgment.

## Abbreviations

DIT – Defining Issues Test

P Score – Post-conventional score response percentage

U score – Utilizer score on the DIT, the extent to which subjects selected responses that were consistent with a reliance on justice-based reasoning.