Journal of Business and Accounting
Vol. 9, No. 1; Fall 2016

CHANGES IN STUDENT MORAL REASONING LEVELS FROM EXPOSURE TO ETHICS INTERVENTIONS IN A BUSINESS SCHOOL CURRICULUM

Lisa Flynn
Howard Buchan
SUNY Oneonta

ABSTRACT: In the wake of corporate scandals, accounting frauds, and losses of billions of dollars in the early 2000s, the Sarbanes-Oxley Act of 2002 (SOX) was enacted to restore investor faith and confidence in the markets and to remediate the wrongdoing seemingly prevalent in corporate America. One of the educational impacts of SOX was a demand for increased and improved teaching of business ethics to students enrolled in collegiate business programs, and this paper assesses the impact of the greater emphasis placed on business ethics instruction. The current study measures moral reasoning scores of business students prior to any specific business ethics instruction and compares those scores to their moral reasoning scores near the conclusion of their educational programs, after several business ethics interventions. The research is conducted using the Defining Issues Test 2, using paired sample t-test statistics. Results show that student scores increased between pretest and posttest, that male students scored more poorly than female students on both the pretest and the posttest, and that male students showed greater improvement in moral reasoning scores from pretest to posttest than female students. The findings suggest that ethics instruction within business school curricula has a positive impact on the moral development of students within business programs.

Key Words: Business Ethics, Defining Issues Test, Sarbanes Oxley, Curriculum

INTRODUCTION AND BACKGROUND

There are typically three options for accomplishing specific instruction in business ethics. Business curricula may mandate a course external to the business school in ethics or business ethics specifically from a course within liberal arts departments, typically philosophy. Alternatively, business programs may offer a business ethics course as a stand-alone course within the business school. Finally, business ethics instruction may be infused throughout several courses within the business curriculum in lieu of a separate course. The method of instruction under study represents the infusion approach, where students receive an introduction to business ethics in a management fundamentals course and a multi-step approach to reasoning through business dilemmas. Business ethics is reinforced using the same approach to solving dilemmas subsequently in marketing, upper level management, and strategic management courses. In an effort to assess the

effectiveness of the infusion approach, the authors surveyed students (pretest) in the management fundamentals course prior to receiving any ethics instruction. The same students were surveyed several semesters later (posttest) at the conclusion of ethics interventions in the upper level management capstone course. Results of the pretest/posttests were then compared to determine the impact of ethics education.

The study of ethics often investigates individuals' reported thought processes or ways of analyzing ethical issues. Actions are in many ways more difficult to capture and analyze, and as such ethics research often encompasses psychological aspects and moral reasoning about issues with ethical content, as opposed to the study of actions taken. Moral reasoning has its roots in the cognitive moral development theories of Kohlberg (1969), who proposed a three level, six stage (two stages at each level) model. Level one, the pre-conventional level, assumes individuals are primarily concerned with rewards and punishment. This level of moral reasoning has been referred as Personal Interest. Individuals reasoning at level two, the conventional level, consider the consequences of behavior in relation to others, and to laws and other codes of conduct. This level of moral reasoning has been referred to as Maintaining Norms. Level three, the post-conventional level, is the highest level and universal truths become a primary focus, often referred to as Principled Moral Reasoning.

James Rest (1986) proposed a four-component model of cognitive moral decision making that includes cognitive moral development as one component of the overall decision making model. Stage one, moral sensitivity, involves recognizing the ethical component of an issue and determining alternative courses of action. Stage two, moral judgment, relates to cognitive moral development and is the stage where alternatives are weighed against an individual's sense of morality and the most appropriate course of action is identified. The next stage, moral motivation (intent), involves placing moral judgments about the appropriate action above other considerations such as practical expediency and requires an individual to assume personal responsibility for outcomes. The final stage, moral character, requires an individual to carry out his or her moral intent despite obstacles and fatigue that may otherwise prevent the ethical action from being implemented (Rest, et al., 1999a). The stages would seem to logically move in a somewhat sequential fashion, although Rest theorizes that the components of the decision making model interact in a complex reciprocal manner. Notably, several models used in general business ethics research incorporate the four-component model (e.g., Jones and Ryan, 1997; Jones, 1991; Ferrel et al., 1989).

Rest (1979) also developed a survey instrument for assessing individual moral reasoning levels. The resultant Defining Issues Test (DIT) was a practical improvement over prior interview-based methods of discerning levels of moral reasoning. The DIT and its updated version, the DIT2, present a series of moral dilemmas with detailed instructions regarding making an action choice. Participants are also required to rank a list of statements as to the level of importance of each statement's main idea to the participant in judging the situation and choosing an action. Answers provided are used to determine level of cognitive moral development, or moral reasoning.

The current study is grounded in Rest's four component model of cognitive moral decision making and looks at the second component, moral judgment. Moral judgment relates to Kohlberg's work on cognitive moral development and the three levels and six stages described briefly above. The DIT-2 is an instrument used to measure level of cognitive moral development as described by Kohlberg, and is the instrument used to capture students' levels of moral development and any changes therein as a result of ethics instruction at the collegiate level.

METHODOLOGY

The sample: Students within the business school of a medium-sized public college were asked to participate in an ethics assessment using the DIT-2. Willing participants took the DIT-2 in the introductory level management course and then took it again in the capstone course of the business program. The total sample consists of 130 matched surveys. Students were mainly business majors (115), with a few accounting majors in the sample (15). The sample was split at very nearly two-thirds male and one-third female respondents, with 87 males and 43 females participating. Average age of the students was 20 for the pretest and 22 for the posttest.

Students were offered the opportunity to participate and participation was voluntary. It was explained that anonymity was not guaranteed, as names were necessary for the initial pairing of posttest with the pretest. However, upon pairing, students were assigned a number and then names were removed. Demographic information related to age, gender, major, and level of completion of the degree were gathered, but upon assignment of a number to each student, anonymity was maintained as the research progressed with the removal of students' names from the completed surveys.

The instrument, the DIT-2: For both iterations of the survey, the students were presented with five moral dilemmas (referred to as "stories") along with a detailed set of instructions. After reading a given scenario, subjects were asked to select an action choice. At the conclusion of each story was a list of twelve issues/questions that seek to gather information about which items were of highest importance in coming to an action choice for the scenario. Respondents rated (on a 1-5 scale) the importance of each issue to the story. Participants then ranked (in terms of importance) the top four issues. Scores are provided by level, or schema (i.e., Personal Interest, Maintaining Norms and Principled) and are based on the participants' ranking of the issues. Each level score represents the relationship between the actual score for each level to the total possible score. The P-score (principled score), which historically has been the most widely reported index in ethics research using the DIT, represents a quantitative measure of relative weight given to Principled moral reasoning. Thus, the higher the P-score, the greater the use of higher level of moral reasoning. The higher the incidence of principled reasoning when assessing the scenarios, the better the cognitive moral development of the individual. The scoring procedure also provides a test for social desirability bias. Subjects with an "M" (meaningless) score equal to or

greater than eight are eliminated from the sample (Rest, 1993). The DIT has been used to investigate the impact of educational interventions. Most recently Christensen, et al.'s (2016) meta-analysis reviews 43 studies that consider the effect size of several factors using the Defining Interest Test. The primary focus of the analysis relates to accounting students and accounting professionals. However, of particular interest is the impact of embedded ethics instruction where the authors found a statistically significant positive relationship with P-scores. Additionally, in the school at which the research was conducted, business majors and accounting majors take all of the same business core, and as such received the same ethics interventions.

The DIT-2 used in this study represents an updated version of the DIT. Rest et al. (1999b) introduced the new instrument, citing several improvements including the elimination of outdated dilemmas, reduction of the number of scenarios under consideration, and improved reliability checks. Improved validity is primarily due to the new N2 index (Rest, et al., 1997) and the reliability checks. Like Rest et al. (1999b), Bebeau and Thoma (2003) found a strong correlation (r=.79) between the two versions of the instrument.

The new N2 index uses both ranking and rating data. One component of the index is nearly identical to the P-score and the other component is based on the difference between average ratings given to lower stage (Personal Interest) items and the higher stage (Principled) items. The composite N2 is the sum of the P-score and the weighted rating data. Rest, et al's (1997) meta-analysis compares the effect size of the P index and the N2 index and shows that the N2 index generally outperforms the P index based on typical validity criteria. Rest et al. (1999a) cite over 400 published articles in assessing the validity of the DIT. Adequate reliability using Cronbach's alpha was found to be in the upper .70s for the P index and low .80s for the new N2 index (p. 92). Thoma and Dong (2014) provide a comprehensive summary of the evidence supporting the validity and reliability of the DIT, and address a number of questions related to the instrument.

Data Collection and Analysis: For all 130 surveys, scores were calculated for each stage (Personal Interest, Maintaining Norms, Principled Reasoning) along with the overall N2 score. Matched pre-and post-test scores were analyzed using paired samples T-tests at every level of moral reasoning. Students were not broken down by major due to the low number of accounting majors in the sample; however, the data were analyzed by gender to evaluate differences in results between male and female students.

RESULTS

A total of 130 usable matched surveys provide the following results of the study. Basic statistical information is provided in the three panels of Table 1.

Table 1a: Mean Responses for full sample by category for pretest and posttest

Category: (N=130)	Pretest Mean	Posttest Mean
Pre-conventional Level (Personal Interest Category)	30.1	27.2
Conventional Level (Maintaining Norms Category)	34.4	34.9
Post-conventional Level (Principled Reasoning Category)	29.0	31.7
N2 (Overall composite score)	27.6	32.1

Table 1b: Mean Responses by category for pretest and posttest – Males

Category: (N=87)	Pretest Mean	Posttest Mean
Pre-conventional Level (Personal Interest Category)	32.3	29.6
Conventional Level (Maintaining Norms Category)	34.4	35.2
Post-conventional Level (Principled Reasoning Category)	25.8	28.7
N2 (Overall composite score)	23.9	28.5

Table 1c: Mean Responses by category for pretest and posttest – Females

Category: (N=43)	Pretest Mean	Posttest Mean
Pre-conventional Level (Personal Interest Category)	25.8	22.3
Conventional Level (Maintaining Norms Category)	34.6	34.2
Post-conventional Level (Principled Reasoning Category)	35.5	37.8
N2 (Overall composite score)	35.0	39.2

The initial comparison looked at the difference in mean scores from pretest to posttest for the entire sample taken as a whole. Mean scores, along with pretest-posttest differences, t-value, and significance level, are provided in Table 2.

Table 2: T-test of Difference in Means from Pretest to Posttest (N=130)

Category:	Pretest	Posttest	Difference	t-value	Signif.
Personal Interest	30.1	27.2	(2.9)	2.487	.014*
Maintain Norms	34.4	34.9	0.5	0.329	.743
Principled Reasoning	29.0	31.7	2.7	2.211	.029*
N2 Score	27.6	32.1	4.5	3.514	.001*

All differences from pretest to posttest were in the anticipated direction. With greater exposure to ethical interventions that teach higher order ethical reasoning, one would expect a drop in responses that indicate a pre-conventional level of cognitive moral development (the Personal Interest category). The difference in mean score of -2.9 is significant at the .05 level. While there is no statistical difference in mean score for the Maintaining Norms category of conventional cognitive moral development, the difference indicates a move in the right direction of greater conventional thinking. It is noteworthy that the increase in mean score of 2.7 for the post-conventional level, or level of principled moral reasoning, is also significant at the .05 level. In other words, while the score at the conventional level for the sample was virtually unchanged, respondents showed a much lower amount of pre-conventional moral reasoning and a correspondingly higher amount of post-conventional, principled moral reasoning. It is likely that the gains were made from respondents initially scoring at the pre-conventional level increasing to the conventional level, and that simultaneously respondents scoring initially with responses at the conventional level improved on the posttest to more responses at the post-conventional level. The difference in N2 score from pretest to posttest is a composite of all levels of moral reasoning. As such, a decrease in answers at the Personal Interest level coupled with an increase in answers at the Principled Reasoning level in the current study led to the significant mean change in N2 score at the .001 level of significance. Analyzing results by gender revealed interesting differences in levels of cognitive moral development between the male and female respondents. Table 3 displays the results.

Table 3: T-test of Difference in Means by Gender (87 Males, 43 Females)

Category:	Male	Female	t – value	Significance
Personal Interest pretest	32.3	25.8	2.929	.005*
Personal Interest posttest	29.6	22.3	3.675	.000*
Maintain Norms pretest	34.4	34.6	0.091	.928
Maintain Norms posttest	35.2	34.2	0.393	.695
Principled Reasoning pretest	25.8	35.5	3.764	.000*
Principled Reasoning posttest	28.7	37.8	3.377	.001*
N2 Score pretest	23.9	35.0	4.275	.000*
N2 Score posttest	28.5	39.2	4.184	.000*

For every level on both the pretest and posttest, with the exception of the Maintaining Norms category, females' scores were more favorable than males' scores. The female respondents had significantly lower mean responses at the pre-conventional level of Personal Interest on both pretest and posttest. They also had significantly higher Principled Reasoning responses at both the pretest and posttest, leading to the higher N2 scores for females on both survey iterations. These findings are consistent with reported results from other studies that females score higher on the DIT than males (Bebeau, 2002; King & Mayhew, 2002; Christensen, et al., 2016).

We also investigated the difference in change from pretest to posttest between males and females to see if there were any observable difference in impact of ethics discussion and training on male versus female students. The results are reported in Table 4.

Table 4: T-test of Difference in Means from Pretest to Posttest by Gender

Category (N=87 Males, 43 Females)	Pretest Mean	Posttest Mean	Difference	Signif.
Personal Interest – Males	32.3	29.6	(2.7)	.064
Personal Interest – Females	25.8	22.3	(3.5)	.111
Maintain Norms – Males	34.4	35.2	0.8	.579
Maintain Norms – Females	34.6	34.2	(0.4)	.860
Principled Reasoning – Males	25.8	28.7	2.9	.044*
Principled Reasoning – Females	35.5	37.8	2.3	.334
N2 – Males	23.9	28.5	4.6	.004*
N2 – Females	35.0	39.2	4.2	.065

When the difference in score from pretest to posttest are examined by gender, only the male subsample produced differences that were significant at the .05 level of confidence. However, if the confidence level is relaxed to the .10 level, three of the four categories for males show significant differences from pretest to posttest, and the N2 score for females would also be significantly different from pretest to posttest. Nevertheless, it is clear from the data that male respondents had larger changes in scores than female respondents.

DISCUSSION

For the entire sample, Principled Reasoning and N2 scores increased, while Personal Interest scores decreased, between the pretest and the posttest. This result provides positive confirmation of the benefit of ethics training throughout the business curriculum. In a two year period, mean scores at the pre-conventional level decreased almost 3 points, mean scores of the post-conventional level increased 2.7 points, and overall N2 scores increased by 4.5 points. Studies have

demonstrated that scores on the DIT increase with age and also as a result of education (Mayhew & King, 2008; King & Mayhew, 2002). Specifically, Mayhew & King (2008) found that N2 scores increased by 4 points in a pretest/posttest design as a result of ethics interventions. Our results are similar, with an overall sample increase of 4.5 points from pretest to posttest after ethics interventions throughout the curriculum. While the increase in the female respondents' N2 scores was 4.2 points, the male respondents' mean N2 scores increased by 4.6 points.

When looking at the scores of male and female respondents, the current study is consistent with many others that females typically score higher on post-conventional (principled) reasoning than males. For all categories excepting Maintaining Norms, for both pretest and posttest responses, female mean scores were significantly different from male scores. Females demonstrate lower levels of pre-conventional moral reasoning, similar levels of conventional moral reasoning, and higher levels of post-conventional moral reasoning. Significant differences remain after administration of the posttest, indicating that males do not tend to "catch up" with females as a result of ethics instruction. This finding would indicate that more needs to be done to increase the level of principled moral reasoning in males than ethics interventions in business curricula. Surveys of corporate fraud indicate that men are more often the perpetrators of frauds and other corporate abuses (Weiss, 2009). That can partially be explained by the larger presence of men vs. women in upper levels of corporate management, but the findings of this study suggest that moral development levels may also have a part in explaining fraud survey results.

Worthy of note is that changes in mean principled reasoning scores (post-conventional level moral reasoning) and mean N2 scores for male respondents were significant. The numerical change was greater for men than women: the male subsample increased by 2.9 and 4.6 points on the Principled score and N2 score respectively, compared to an increase of 2.3 and 4.2 respectively for the female subsample. The finding that men increased levels of moral reasoning from pretest to posttest is encouraging because even though male mean scores remain significantly lower than female scores at the posttest, men appear to demonstrate greater increases in principled reasoning and N2 scores. Ethics interventions may not be bringing male responses up to the level of female responses, but it may be the case that ethics training throughout a business curriculum has a larger impact on men than women. In other words, though the men did not catch up to the women, it appears that they made greater strides in moral reasoning as a result of ethics interventions. It can certainly be suggested that more ethics training is needed, but the current study provides encouraging evidence that ethics training has a positive impact on students' moral reasoning levels and abilities.

Limitations: The study is limited in its ability to separate increases due to age and college education in general from the impact of ethics interventions. The increases after ethics instruction are similar to increases found by other researchers employing a similar study design and using ethics interventions, providing confirmatory evidence. Additionally, to the extent that the individuals

participating are not representative of students enrolled in business programs at four-year institutions, the results are potentially of low generalizability. We have no specific reasons to conclude that the students surveyed are not representative of other business students. Once again, the similarity in results from earlier studies (Mayhew & King, 2008) provides confidence of the applicability and generalizability of the findings.

CONCLUSION

The current study serves as confirmatory evidence of previous studies using the DIT-2. It also advances the field of business ethics training by suggesting that while men remain at lower levels of higher order (principled) moral reasoning than females after ethics instruction, men benefit more greatly from that ethics instruction, as demonstrated by significant differences in mean scores from pretest to posttest for men but not for women. While more needs to be done to further the ethical development of business students and effect changes in behavior of business professionals, the current study suggests that ethics instruction as part of a business curriculum serves to increase moral reasoning levels in business students.

REFERENCES

Bebeau, M.J. (2002). The Defining Issues Test and the Four Component Model. Journal of Moral Education, 31(3), 271-295.

Bebeau, M.J. & Thoma, S.J. (2003). Guide for DIT-2. MN: University of Minnesota.

Christensen, A.L., Cote, J. & Latham, C.K. (2016). Insights regarding the applicability of the Defining Issues Test to advance ethics research with accounting students: A meta-analytic review. Journal of Business Ethics, 133(1), 141-163.

Ferrel, O., Gresham, L., & J. Fraedrich, J. (1989). A synthesis of ethical decision models for marketing. Journal of Macromarketing, 55-64.

Jones, T.M. (1991). Ethical decision making by individuals in organizations: An issue contingent model. Academy of Management Review 16(2), 366-395.

Jones, T.M. & Ryan, L.V. (1997). The link between ethical judgment and action in organizations: A moral approbation model. Organization Science 8(6): 663-680.

King, P. & Mayhew, M.J. (2002). Moral Judgment in higher education: Insights from the Defining Issues Test. Journal of Moral Education, 31(3), 247-270.

Kohlberg, L. (1969). Stages and sequences: The cognitive developmental approach to socialization. In D. Goslin (ed.), Handbook of Socialization Theory and Research. Chicago: Rand McNally.

Mayhew, M.J. & King, P. (2008). How curricular content and pedagogical strategies affect moral reasoning development in college students. Journal of Moral Education, 37(1), 17-40.

Rest, J. (1979). Development in Judging Moral Issues. Minneapolis: University of Minnesota Press.

Rest, J. (1986). Moral Development: Advances in Research and Theory. New York: Praeger.

Rest, J. (1993). Guide for the Defining Issues Test-How to Use the Optical Scan Forms and the Center's Scoring Service. University of Minnesota: Center for the Study of Ethical Development.

Rest, J. (1994). Background: theory and research. In James R. Rest & Darcia Narvaez (eds.) Moral Development in the Profession. Hillsdale: Lawrence Erlbaum Associates.

Rest, J., Thoma, S.J., Narvaez, D., & Bebeau, M.J. (1997). Alchemy and beyond: Indexing the Defining Issues Test. Journal of Educational Psychology, 89(3), 498-507.

Rest, J., Narvaez, D., Bebeau, M.J. & Thoma, S.J. (1999a). Postconvential Moral Reasoning: A Neo-Kohlbergian Approach. Mahwah, NJ: Lawrence Erlbaum Associates.

Rest, J., Narvaez, D., Bebeau, M.J. & Thoma, S.J. (1999b). Devising and testing a revised instrument of moral judgment. Journal of Educational Psychology, 91(4), 644-659.

Thoma, J. & Dong, Y. (2014). The defining issues test of moral judgment development. American Psychological Association-Behavioral Development Bulletin, 19(3), 55-61.

Weiss, J. W. (2009). Business Ethics: A Stakeholder and Issues Management Approach, 5th ed. Mason, OH: South-Western Cengage.

ACKNOWLEDGMENT: The authors wish to thank the participants of the Annual American Society of Business and Behavioral Sciences conferences who provided comments and suggestions on earlier drafts of the paper.