

The Routes of Moral Development and the Impact of Exposure to the Milgram Obedience Study

Jerry Paul Sheppard
Marnie Young

ABSTRACT. This article examines how business students route themselves through the process of cognitive moral development (CMD) to arrive at a more autonomous level of CMD when there is an impetus to do so. In this study, two groups were given Rest's Defining Issues Test; half the test 1 week and half three weeks later. In between, one group viewed a film of Milgram's obedience study as a stimulus towards a more autonomous level of CMD. The results of the analysis indicate that viewing the Milgram study produced a positive response regarding subjects' level of autonomous CMD. However, the response was not uniform across the subject pool. Females showed a greater consistent significant positive response to viewing Milgram while male subjects varied their response contingent upon their functional area of study. While subjects' functional area of study alone made little difference in the results, when

taken in conjunction with gender, significant differences were found between groups. Thus, researchers should take care when investigating differences between subjects' area of study since gender differences may be present even within an apparently homogenous population-like business students.

KEY WORDS: cognitive moral development, DIT, Kohlberg, Milgram, obedience, Rest

Marnie Young is currently completing her Masters in Counseling Psychology at the Adler School of Professional Psychology in Vancouver, BC, Canada. This article was developed from Her Honours BA thesis from Laurentian University where she graduated Cum Laude in 2000. She has presented research papers at the Administrative Sciences Association of Canada and the Academy of Management. Her paper presented to the Administrative Sciences Association of Canada in 2000 won Honourable Mention in the Business Education Division.

Jerry Paul Sheppard is an Associate Professor Strategic Management at Simon Fraser University, Burnaby, BC, Canada. His research focuses primarily on organizational decline, turnaround, survival and failure. His PhD dissertation from the University of Washington in 1989 won the Strategic Management Society Best Dissertation Award. He has published Journal articles in the Journal of Management, Long Range Planning and Social Science Research. His most recent work includes Strategic Management: Competitiveness & Globalization, 2nd Canadian Edition, with Hitt, Ireland, Hoskisson, and Rowe.

Hear the names Enron and World Com and what comes to mind? "Phoney earnings, inflated revenues...a crisis of investor confidence the likes of which hasn't been seen since...the Great Depression" (Nocera, 2002: 65). How did this happen? CEO's like Enron's Jeff Skilling and World Com's Bernie Ebbers were not lone mavericks whose rogue actions bankrupted their firms. They enlisted teams of managers and accountants to hide their questionable acts (Behr and Witt, 2005; Moberg and Romar, 2005).

Were those who helped craft the accounting slight-of-hand at these firms just following orders? While Vince Kaminski helped Enron hide billions in debt, Cynthia Cooper at WorldCom fought with her bosses to pursue the largest fraud case in U.S. history. How do we get more of the latter type of manager? How do we get managers to take the moral high road despite orders by higher authority to do otherwise? One place to start is to show future managers the pitfalls of obedience to authority. In this study, we showed students in one section of a third year undergraduate business ethics class the classic 45-minute film, *Obedience to Authority* (Milgram, 1965) and compared the change in their level of cogitative moral development to those in another section of the same class who did not see the film.

To address the question, “would viewing the Milgram film aid in students’ cognitive moral development?” we realized that a response to the film would require students to develop greater independent thinking. To study the change in such independent cognitive moral development (CMD) we took a closer look inside Kohlberg’s (1969) model of CMD to see how individuals progress through moral development substages when faced with a new ethical issue. Also, we sought to uncover what relevance gender¹ and functional area of study had on responses. To clarify what we mean by CMD and obedience to authority, some background is in order.

Theoretical background

A wide range of influences can impact one’s moral judgement. Our principle concern here is how learning about obedience to authority can improve moral judgement by raising one’s level of autonomous CMD.

Cognitive moral judgement and CMD

Rest (1986), asserts that there are four component processes involved in taking moral action: (1) recognize the moral issue: interpret the situation and identify the moral problem; (2) make the moral judgement: figure out what ought to be done and

devise a plan of action to apply the relevant moral standard (e.g. judge which one course of action is morally right); (3) establish moral intent: evaluate how various courses of action serve moral values and decide which action one will actually try to pursue; (4) execute and implement the moral plan of action: figure out the concrete steps involved in executing the plan and have the ego strength and self-regulatory processes needed to follow through with it.

Rest’s model assumes moral action is the product of each of the four processes operating in combination (like stones forming an arch). A failure to act morally is due to a breakdown in one or more components. Thoma, Rest and Davison (1991) note that in executing the second component – making the moral judgement – individuals may rely on notions of justice (Kohlberg, 1969), social norms (Nisan, 1984; Nucci, 1982) or religious prescription (Kennedy and Lawton, 1998) to construct an ideal action choice. To the extent to which an individual is driven to make the moral judgement based on a justice perspective, Kohlbergian-based notions of moral judgement processes become important. Thus, an individual’s level of CMD plays an important role in applying a relevant moral standard. Of particular importance in the Kohlberg model are Kohlberg’s six hierarchical stages of CMD as shown in Figure 1.

While Kohlberg’s model is well accepted, it is less than perfect in presenting an orderly set of stages. Some research shows that when faced with new

		Level I: Preconventional Level Focus on the Individual		Level II: Conventional Level Focus on the Group		Level III: Postconventional Level Focus on the Inner Self	
		Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6
		Self Centered Ethics	Internalized Shared Norms	Ethics of Conformity	Soc. System / Conscience Maintenance	Social Contract & Prior Rights	Universal Ethical Principles
Substage A: Heteronomous	One obeys rules to avoid punishment or gain reward (the ethics of convenience).	One recognizes the need to accommodate the interests of others.	One is guided by what is accepted by others; One seeks group approval & has group loyalty.	One obeys given professional codes, civic and religious laws as well as the commonly accepted roles and expectations.	One is drawn to Social Contract theory/ ethics. Conduct is the result of consensus from due process.		
			(Rules may be heteronomously accepted or one may make an autonomous conclusion that the rules are moral.)	(Rules may be heteronomously accepted or one autonomously concludes that the rules are moral.)	Changes in the law are considered for socially useful purposes or maximum utility.		
Substage B: Autonomous							One follows unwritten global ethical values from theories of justice, duties & human rights.
		Stage 1	Stage 2	Stage 3	Stage 4	Stage 5	Stage 6

Figure 1. Kohlberg’s moral development stages.

moral dilemmas, people revert back to the use of lower moral development stages (Reed, 1997). To address this inconsistency, Kohlberg created substages in the model (Kohlberg, 1981). Known as the heteronomous A substage and autonomous B substage, these substages occur in at least stages 3–5 (Reed, 1997). B substage reasoning differs from A substage reasoning due to its inclination for independent choice in a dilemma and some formal features of reasoning. “Norms and considerations of consequences are brought to bear on a dilemma choice in B substage reasoning, whatever the justice stage, such that the autonomous choice is preferred” Reed (1997: 80–81).

Thus, one may chart various routes through Kohlberg’s substages. Individuals may follow direct routes through stages 1, 2, 3A, 3B, 4A, 4B, 5B and 6 for example, or chart a more circuitous route (like 1, 2, 3A, 4A, 3B, 4B, 5B and 6)² – i.e. some people proceed more directly than others in developing better moral judgements. Consistent with Nelson’s (2000) call for specific stage analyses and stage pattern

examination, we took a closer look at how different people route themselves through the substages.

The Rest (1986) and Kohlberg (1969) models are imbedded within larger forces (Jones, 1991; Trevino, 1986) (see Figure 2). Trevino’s (1986) interactionist model contained two moderators of moral behaviour: individual and situational variables. Trevino lists ego strength (the power of one’s self identity), field independence (the ability to focus on the relevant elements of a problem) and locus of control (the extent to which one feels that they, not others, control events) as salient dimensions interacting with an individual’s morals.³ Situational variables that affect moral behaviour (Trevino, 1986; White, 2002) are: work characteristics (e.g. the frequency of moral dilemmas on the job), organizational culture (e.g. the degree of organizational member autonomy) and immediate job context (e.g. organizational rewards and punishments). Thus, the job itself or a predisposition on the part of those occupying it may impact ones’ moral decisions.⁴ Job impacts call for study of what tasks the job involves that result in

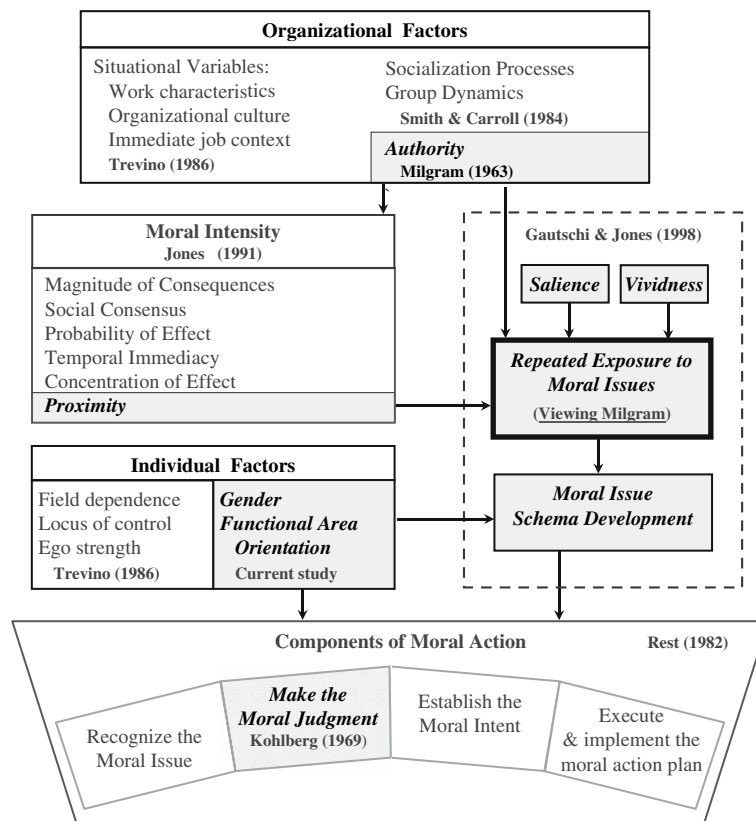


Figure 2. Theoretical impacts from exposure to the Milgram video.

more, or less, moral decisions.⁵ Individual impacts, as studied here, call for study of who is filling the job.

Building on Trevino's work, Jones (1991) – suggested that one's displayed stage of CMD may be contingent upon the issues involved in the moral dilemma – or what he calls *moral intensity*. For Jones, *moral intensity*, has six components:

1. Magnitude of consequences – greater harm to greater numbers is of greater consequence;
2. Social consensus – the extent to which society agrees that an act is right or wrong;
3. Probability of effect – the degree to which harm is likely;
4. Temporal immediacy – the length of time it takes to have an impact;
5. Concentration of effect – e.g. cheating 10 people out of \$10,000 versus 10,000 out of \$10;
6. Proximity – the physical or psychological closeness or distance from those impacted.

Jones, citing Smith and Carroll (1984), notes three important organizational factors that influence the chance of engaging in moral behaviour: group dynamics (e.g. groupthink), authority factors (see below) and socialization processes (e.g. corporate culture). The current research looks at the influence that learning about authority factors will have on moral judgement and specifically ones' level of CMD.

By using Milgram as a stimulus, we have subjects deal with an issue (obedience) that is salient and vivid. "A salient moral issue stands out from other issues" (Fiske and Taylor, 1984: 180). A vivid stimulus is "(a) emotionally interesting, (b) concrete and imagery-provoking, and (c) proximate in a sensory, temporal or spatial way" (Nisbett and Ross, 1980: 45). Salience and vividness "...affect recognition, which in turn affects schema development" (Gautschi and Jones, 1998: 208).⁶ Using this idea of schema development, Gautschi and Jones found taking a business ethics class helped students develop schemas that allow them to better identify moral dilemmas – the first of Rest's four components.

CMD and Milgram's Study on obedience to authority

Since business-people's actions typically occur within hierarchies, addressing the behaviour of the

individual in an organization is critically important. Learning about obedience to authority may create more morally developed individuals. This may be a vital step towards reducing future management abuses of the type discussed earlier.

In the early 1960s, Stanley Milgram conducted a series of experiments in which naïve subjects – in the guise of a learning experiment – gave seemingly painful electric shocks to an accomplice of the experimenter (Milgram, 1974). The study tested how obediently subjects would follow the researcher's commands. Milgram varied the proximity of both the researcher and the victim to the naïve subject. He found that the degree of obedience varied with, what Milgram called, the psychological distance – what Jones (1991) calls proximity – between the victim, the researcher as an authority, and the naïve subject.

Miceli (1996) notes the connection between Milgram and morality in organizations: "...[P]eople Identify with their duties, and view them as coming from higher authority. This leads people to view the duties they are given as legitimate and reduces the chance of orders being questioned..." (Miceli, 1996: 704). Thus, in the presence of authority, amoral decisions become a more likely possibility.

Milgram's film footage repeatedly shows subjects coming to grips with a declared need to go further in the experiment with higher shock voltages. The Milgram film is thus a repeated exposure to a moral issue which is a salient and vivid stimulus that addresses issues of moral intensity (proximity) and organizational factors (authority).⁷ This exposure to moral issues creates a chance for subjects to develop moral schema and address moral judgements employing a more postconventional level of moral development.⁸ These relationships are shown in Figure 2.

Films have been employed in studies to try to improve students' level of CMD (Schlaefli et al. 1985). Oberlander K. J. (unpublished) employed a 50-minute film and a 50-minute discussion of moral dilemmas with no significant effect. However, Goddard R. C. (unpublished), using films about educational principles of person centred therapy and principles of assertiveness, obtained significant change results. Self et al. (1993) observed significant improvement in students' CMD for a film course in ethics. The effects they found were comparable to those obtained with

lecture and case-studies (King and Mayhew, 2002).⁹ Thus, the null hypothesis is that:

H1_N: Viewing Milgram will have no impact on subjects' level of CMD.

Viewing Milgram, gender and CMD

Observed gender differences in CMD may be the result of a male-biased interview method of moral assessments (Holstein, 1976; Rest, 1979: 120). Gilligan (1977, 1982) and Gilligan et al. (1988) however, note that there really are gender-based differences in how men and women resolve moral conflicts. Gilligan notes that Kohlberg's stages are focused on a justice view of morality, but there is a "care" perspective – prevalent among women – that should be studied as well. "A justice perspective draws attention to problems of inequality and oppression and holds up an ideal of reciprocity and equal respect. A care perspective draws attention to problems of detachment or abandonment and holds up an ideal of attention and response to need" Gilligan and Attanucci (1988: 73).

Yet, the "justice" versus "care" perspective may not discriminate well between men and women.¹⁰ Axelrod D. E. (unpublished) found both justice and care moral orientations were demonstrated in her sample of business students and that their moral orientations were not associated with the subjects' gender.

There are numerous studies showing little difference between men and women in levels of CMD.¹¹ Weber and Wasieleski (2001), note that Derry's (1987, 1989) tests of Gilligan's research found no conclusive differences between moral reasoning orientations of females and males. Walker's (1984) meta-studies of prior Kohlberg type work "...supported the conclusion that the overall pattern is one of non-significant gender differences in moral reasoning" (Walker, 1984: 677).

Some studies support the idea that women have greater concern for ethical issues than men. St. Pierre et al. (1990) found that female students had significantly higher levels of CMD. Beltramini et al. (1984) and Peterson et al. (1991) in surveys of thousands of students found that female business students were more concerned with ethical issues in

business than males. In a meta-study by Borkowski and Ugras (1998), 47 gender-related studies showed females were more, the same, or less ethical than males, depending on the study.

Regarding interventions undertaken to improve subjects' moral judgement or level of CMD (as is being done here with Milgram), the results are mixed. A meta-study by Bebeau (2002) showed that interventions resulted in stronger responses by females. Interventions by Beard (1983) and Hoffinan D. A. (unpublished) showed no gender difference in the degree of CMD change caused by the intervention. Based on this range of results, the following null hypothesis is that:

H2_N: There will be no significant difference in the impact Milgram has on female subjects' level of CMD versus male subjects' level of CMD.

Viewing Milgram, functional area background and CMD

In addition to gender differences, Rest (1994) found that there was a wide range of differences among both students and those practicing in various professions. Rest (1994) compiled *P*-score data on thousands of subjects. *P*-scores represent the percentage of Kohlberg's postconventional morality stages that subjects employ to make their decisions. Rest's compilation shows that *P*-scores for groups range from a 21.9 for Junior High School students, to 65.2 for Philosophy and Political Science Graduate students. In between those extremes lie College students (42.3), Business Graduate students (42.7), Physicians (49.2), Law Students (52.2) and Liberal Seminarians (59.8), among others.

In a meta-study by Borkowski and Ugras (1998), the majority of 30 studies looking at undergraduate majors showed no significant differences. Weber and Wasieleski (2001) found managers whose work tasks are mainly policy and strategy (what they labelled "supervisory") employed a higher level of moral reasoning than "technical" managers (managers in sales, engineering, finance, accounting and information technology).¹² Accounting and Finance students may be a relatively different population from the rest of the business school. Sneed and Morgan (1999) found that those entering the

accounting curriculum had higher quantitative skills and lower problem solving skills than those in other areas of the business faculty. Pritchard et al. (2004) found that students majoring in accounting and finance had higher basic algebra scores than management and marketing majors. Given the use of these divisions in these studies, separating out accounting and finance majors are a logical step.

Those in more quantitative-oriented areas may show lower degrees of change in CMD than people in less number-driven areas. This difference is usually referred to in discussions of situations where daily personnel management, problem solving skills and leadership (including, logically, moral leadership) are found to be more critical than analytical skills (e.g. Greising, 1989, Rifkin, 1996, or Hobel, 2002). Jeffery (1993) found that less quantitative education is useful for quantitative-orientated students – accounting majors tended to have higher levels of CMD growth during the course of their university career if they attended a liberal arts school. In the current study, we grouped those in quantitative areas (accounting, economics and finance) into one group and all others into a second group.¹³ The null hypothesis here is that:

H3_N: There will be no significant difference in the impact Milgram has on quantitative-oriented subjects' cognitive moral development versus non-quantitative-oriented subjects' CMD.

Viewing Milgram, gender, functional area of study and CMD

Given the above background, we may logically expect that there would be compound impacts of viewing Milgram, gender and functional area of study on CMD. In a study of the compound impacts of gender and seniority within a profession, Bernardi and Arnold (1997) found that, as time went on, more ethical females and less ethical males tended to stay in the accounting profession. Thus, for the compound impacts of Milgram, gender and functional area on CMD, the null hypothesis is:

H4_N: There will be no significant difference in the impact Milgram has on subjects' CMD based on different functional areas of study and gender.

Method

There are a number of methods that can be employed to measure CMD.¹⁴ We employed Rest's Defining Issues Test (DIT) for several reasons. First, it is well recognized and accepted. Evans P. K. S. (unpublished) notes that, "the DIT is the most widely used objective and reliable psychometric instrument in the field of moral psychology" (Evans, P. K. S., unpublished). Second, the DIT's reliability, validity and stability are well established (White, 2002). Third, the DIT is a measure of moral reasoning that "assesses an individual's propensity to use concepts of justice based on social cooperation" (Elm and Weber, 1994: 350). This is important for the current study in that such a measure allows for inclusion of notions of social cooperation or, conversely, given our concern, independence. Fourth, the DIT, allows subjects' responses to be quickly and mechanically scored; a method like Kohlberg's Moral Judgment Interview, for example, typically requires at least seven distinct manual steps for scoring (Weber, 1991). Last, though not unique to the DIT, scores of Milgram viewers and non-viewers can be compared using straightforward statistical methods: ANOVAs and *T*-tests. Rest (1994) lists these as appropriate statistical measures for the DIT.¹⁵

The DIT consists of six moral decision scenarios. After each scenario, subjects are asked to arrive at a decision (e.g. steal medicine to save a dying spouse). A total of 12 statements addressing issues in that scenario follow. Subjects rank the top four statements that helped them arrive at their decision. Of the 12 statements, three to four address thinking at Stage 5 or 6: the postconventional morality stages. If the top ranked statement addresses a postconventional morality stage it receives four points, if the second most important ranked statement addresses a postconventional morality stage it receives three points and so on. Between the six scenarios, there are 27 statements employing Stage 5 or 6 reasoning – resulting in a maximum raw score of 57.¹⁶ To arrive at an approximate percentage score (and for ease of calculation and consistency), researchers divide the raw score by 0.6 to create a *P*-score – the maximum *P*-score is thus 95 (95 = 57/0.6). For the current research, the DIT was split in half, each half containing three scenarios.¹⁷ Part 1 of the test contained a maximum *P*-score of 28 and Part 2 a maximum

score of 29. To make the scores of each part uniform, the raw scores in each part were transformed to have a maximum of 95 $((\text{Part 1 raw score}/28) \times 95)$ and $(\text{Part 2 raw score}/29) \times 95)$.

To investigate if participants were moving towards more independent moral judgements we broke the *P*-score down into heteronomous components (stage 5A) and autonomous components (stages 5B+6, or “5B+” hereafter).¹⁸ Each score (5A and 5B+) was adjusted to account for possible different maximum scores in each section.¹⁹ Thus, the dependent variables in this study are the changes in subjects’ *P*-score, as well as its 5A and 5B+ components, as measured by Rest’s (1994) DIT. Given that viewing Milgram should discourage heteronomous views (5A) and promote autonomous ones (5B+), there are four possible outcomes to viewing Milgram:²⁰

1. No change to 5A and 5B+ scores (what we call a *Retrenchment* to prior stages);
2. No change to 5A and a significant increase in 5B+ (what we call *Principle Addition*);
3. Significant decrease in 5A and no change to 5B+ (what we call *Rule Response* to prior stages);
4. Significant decrease in 5A and a significant increase in 5B+ (what we call *Principle Substitution*).

These patterns are shown graphically in Figure 3. Thus, in addition to a change in *P*-score, we were interested in instances where 5B+ scores increase: *Principle Addition* and *Principle Substitution* types.

Also included in the DIT are several meaningless statements. If a subject picks these phrases as important, it is an indicator that they were not certain of the questions being asked. Rest (1994) notes that subjects who score more than eight points on meaningless questions should be removed from the subject pool.²¹

Watching Milgram and taking the DIT can be time consuming and may limit the number of willing participants. Uddin and Gillett’s (2002) study of Chief Financial Officers involved taking the DIT and other tests less time consuming than viewing Milgram. Their study had a 6 percent usable return ratio (2000 requests, 122 completed). Similar return rates might be expected for the current study if we attempted to obtain a large sample of managers; as such, there would likely be serious questions about response bias and validity of the results. Also, since this study involves measuring subjects’ change in CMD, controlling for influences other than Milgram between testing sessions was also important. Studying a diverse group of managers may make such control difficult; scheduling the timing of the tests relative to viewing Milgram may have to be varied

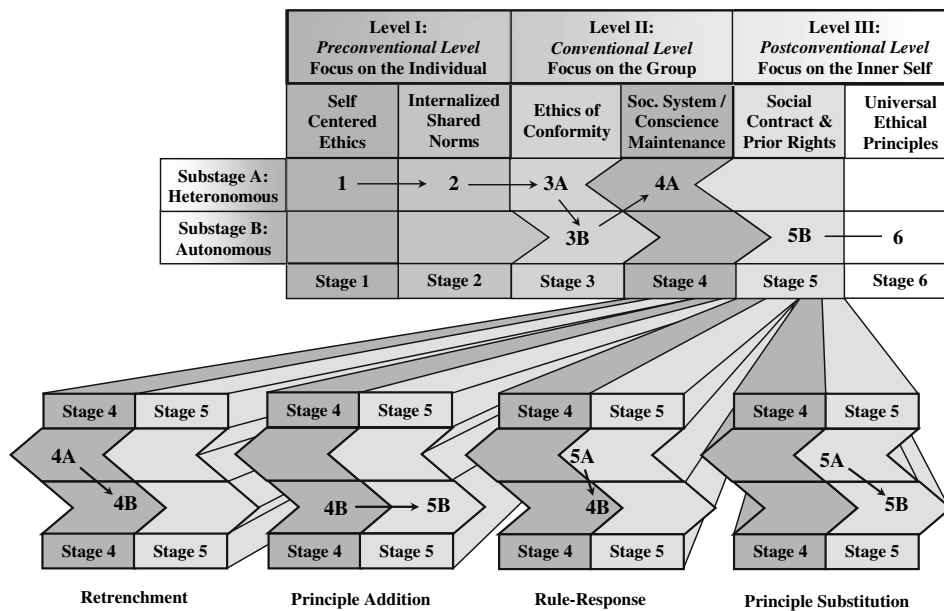


Figure 3. Four types of progression through Kohlberg’s moral development substages

due to professionals' busy schedules and so may call the results into question.

To overcome the above problems, we employed a sample of business students. Since Rest (1994) noted a consistent positive relationship between education and moral judgement, one might conclude that students have lower CMD scores than managers. Yet, studies by Wimalasiri et al. (1996) and Wimalasiri (2001) showed no significant difference in the level of CMD between managers and business students.²² Also, an advantage to using students is that it allows us to identify individual, not organizational, influences; subjects have self-selected the profession rather than being selected by an organization. Thus, these subjects represent individual predispositions towards patterns of CMD and not organizational preferences. To insure we were observing effects from the educational event and not organizational affects from work-related experience, we kept only subjects with limited work experience and culled those who had worked full time for more than 3 years.²³

Procedure

Subjects in different class sections of a *Business, Ethics and Society* course were asked to fill out a questionnaire and three scenarios from the Defining Issues Test (DIT Part 1) at the beginning of the first class of the term. The questionnaire inquired about age, gender, major, university credits earned and business credits taken.

In the last part of the second week of class, the 45-minute film *Obedience to Authority* (Milgram, 1965), was shown to one of the two classes. The majority of documentary film footage shown to students in this study was from a variation of the experiment, where Milgram created a high degree of psychological distance from the victim and low degree of distance to the researcher. In this situation, where the researcher was seated next to the subject and the subject had audio contact with the victim in another room, obedience was higher than in most variations. The film shows the in-depth progress and comments of six male subjects – some of whom went to the highest voltage and others that did not. Milgram also briefly shows variations where

psychological distance from the victim and researcher was altered – with a resultant change in the degree of obedience.²⁴

In the beginning of the fourth class, both classes filled out a questionnaire and three more scenarios from the Defining Issues Test (DIT Part 2). At the end of the DIT Part 2, subjects were asked on the form if they had seen the Milgram Obedience video. This was done to ensure that the student was present on the day of the film and had stayed in class to see it.

Limitations

The DIT has two limitations in measuring moral development. First, the DIT does not measure the care perspective. However, as discussed earlier, one may rely on notions of justice as a principal component in making a moral judgement and the DIT provides some measure of one's ability to make that moral judgement through measurement of their level of cognitive moral development. Second, the DIT is limited in its ability to differentiate between A and B substages and can only do so at Stage 5. This limits our ability to tell if subjects who have a decline in 5A have an increase in 4B scores. In spite of this, we employed the DIT due to its reliability, widespread acceptance and ability to allow us to draw some preliminary conclusions regarding subjects' movement within postconventional morality stages.

There are two other caveats to this study: (1) the snapshot nature of the study and (2) the related possibility of losing short-term moral developmental progress subjects may make. Regarding the snapshot nature of the study, Rest (1986) warns that short-term learning exercises such as the one here seldom yield significant results. It is more typical to test students at the beginning and end of a course lasting months (still, this is two snapshots, albeit farther apart than what was done here). To isolate the impact of the Milgram film however, the current study was limited to taking such a snapshot. However, breaking down the measurement into 5A and 5B+ should aid in the detection of more subtle short-term changes. Is there a possibility of losing short-term cognitive moral developmental progress if it is not reinforced? There is a tradition of viewing CMD as a developmental skill akin to walking or talking

(Kohlberg, 1981). While there are circumstances that one could imagine which might cause such developed skills to atrophy,²⁵ it is unlikely that once gaining a developmental skill one would lose it.

Results

We selected subjects from a volunteer pool of 180 students in two sections of an undergraduate *Business, Ethics and Society* course at a major North American university. Of these, 140 completed the DIT Part 1 and 100 subjects completed Part 2. Of these, 84 Part 1 surveys could be matched to the same person taking Part 2. Five surveys were discarded due to the subjects scoring more than eight on meaningless questions (Rest, 1994). Five surveys were discarded due to subjects having more than limited work experience.²⁶ The remaining 74 subjects ranged in age from 20 to 26 years old; the average being 22.

There were no statistically significant differences in the baseline (DIT Part 1) scores for 5A, 5B+ or *P*-scores (see Appendix 1). Thus, those who later viewed Milgram did not initially score differently from those who did not later see the film.

Viewing Milgram and CMD

There were some significant changes that occurred between Part-1 and Part-2 of the DIT. The Part 1/Part 2 paired *T*-tests included in Appendix 1 show that for both viewers and non-viewers, a significant ($p < 0.05$) drop in Stage 5A reasoning and a significant increase in 5B+ reasoning occurred. The changes for viewers, however, (Part 1/Part 2 paired *T*-tests of -5.90 for 5A and 7.29 for 5B+) were more pronounced than for non-viewers (Part 1/Part 2 paired *T*-tests of -2.35 for 5A and 3.50 for 5B+).

Those who did not view Milgram reduced their use of Stage 5A to a significantly greater degree than those who did. The viewing group increased use of Stage 5B+ reasoning to a significantly greater degree than the non-viewers. Yet, exposure to Milgram did not lead to a significant change in subject *P*-scores.²⁷ For the whole sample, both viewers and non-

viewers employed *Principle Substitution* (see Table I). Those viewing Milgram had a statistically greater change in 5A and 5B+ scores than non-viewers.

Viewing Milgram, gender and CMD

To investigate pair-wise differences between groups a Tukey Procedure was employed. Tukey is most appropriate when – as in the current study – exploratory investigation into pair-wise comparisons is sought (Neter et al., 1996). Also, Tukey performs better in reducing the chance of error for incorrectly accepting a false hypothesis as true (Cherry, 2000). In the appendices, all scores with the notation “L” for Low in the Tukey Procedure are statistically significantly ($p < 0.05$) lower than all scores labelled with the notation “H” for High (and “H”s are significantly different from “L”s).

Defining Issues Test Part 2 showed that males who viewed Milgram used Stage 5A reasoning the least. Males who did not see the video used Stage 5A reasoning the most. All groups (except for male non-viewers) showed significant declines ($p < 0.05$) in use of Stage 5A reasoning. Among viewers, *T*-tests showed females had statistically higher Part 2, 5A and *P*-scores than males (shown as “Viewers T” in Appendix 1). The *F*-score of 2.58 for Part 2 Stage 5B+ shows some difference between groups (but only at $p < 0.06$). The Tukey Procedure showed no significant pair-wise differences between groups in use of Stage 5B+ reasoning in Part 2.

All groups showed significant ($p < 0.05$) increases in their use of Stage 5B+ reasoning given the Part 1/Part 2 paired *T*-tests. The *F*-score of 3.03 indicates a significant difference between groups in their change in Stage 5B+ reasoning. The Tukey Procedure showed no significant pair-wise differences between groups in their change in use of Stage 5B+ reasoning. Finally, overall Part 2 *P*-scores given a Tukey Procedure showed that males who viewed Milgram scored significantly lower than females who viewed Milgram. There were no significant changes found between *P*-scores in the DIT Part 1 versus DIT Part 2 paired *T*-tests (see Appendix 1). Among gender groupings, males who did not view the film were the only ones not employing *Principle Substitution* and instead employed *Principle Addition*.

TABLE I
Male/female, quantitative/non-quantitative functional area and Milgram results summary

		Milgram Film	n	T-Tests		Change Type	Additional Comments	
				5A	5B+			
Full Sample	{	Viewed	39	Sig.↓	Sig.↑	Principle Substitution	↑	
		Unviewed	35	Sig.↓	Sig.↑	Principle Substitution	↑	
Break-down by Gender	Male {	Viewed	12	Sig.↓	Sig.↑	Principle Substitution	↑	
		Unviewed	21		Sig.↑	Principle Addition	↔	
	Female {	Viewed	27	Sig.↓	Sig.↑	Principle Substitution	↑	
		Unviewed	14	Sig.↓	Sig.↑	Principle Substitution	↑	
Break-down by Functional area of study	Quant. {	Viewed	22	Sig.↓	Sig.↑	Principle Substitution	↑	
		Unviewed	17			Retrenchment	○	
	Non-Quant. {	Viewed	17	Sig.↓	Sig.↑	Principle Substitution	↑	
		Unviewed	18		Sig.↑	Principle Addition	↔	
Break-down for all Eight Groups	♂ {	Quant. {	Viewed	6		Sig.↑	Principle Addition	↔
			Unviewed	11			Retrenchment	○
		Non-Quant. {	Viewed	6	Sig.↓		Rule Response	↓
			Unviewed	10		Sig.↑	Principle Addition	↔
	♀ {	Quant. {	Viewed	16	Sig.↓	Sig.↑	Principle Substitution	↑
			Unviewed	6			Retrenchment	○
		Non-Quant. {	Viewed	11	Sig.↓	Sig.↑	Principle Substitution	↑
			Unviewed	8	Sig.↓		Rule Response	↓

Viewing Milgram, functional area of study and CMD

When functional area of study and the impact of Milgram are considered there were no between-group differences for Part 2 DIT Stage 5A, 5B+ or P-Score. However, all groups, except quantitative non-viewers, showed a significant ($p < 0.05$) decrease in use of Stage 5A and a significant increase ($p < 0.05$) in use of Stage 5B+ given the results of the Part 1/Part 2 paired *T*-tests (see Appendix 1).

The *F*-score of 3.17 for change in Stage 5B+ indicates a significant difference between groups – quantitative viewers scoring highest. The Tukey Procedure shows that there were no significant pairwise differences between groups in any of the scores. Milgram had some impact with quantitative subjects to the extent that this group recorded the highest change in their use of Stage 5B+. All subjects who viewed the Milgram film employed *Principle Substitution*; those who did not used *Retrenchment* (quantitative non-viewers) or *Principle Addition* (non-quantitative non viewers).

Viewing Milgram, gender, functional area of study and CMD

When all eight groups were studied, the Tukey procedure results on the DIT Part 2 scores (see Appendix 2), show that the most extreme 5A differences occurred with male non-quantitative subjects. Among this group, viewers scored the lowest of all groups in their use of Part 2 Stage 5A reasoning, non-viewers the highest (i.e., in the end, males who did not view Milgram had the highest use of Stage 5A reasoning). The Part 1/Part 2 paired *T*-tests show that non-quantitative viewers – male and female – had significant declines in their use of Stage 5A.

For Stage 5B+ the Tukey procedure for DIT Part 2 showed a significant ($p < 0.05$) difference between quantitative male non-viewers (lowest) and quantitative females viewers (highest). Females, with the exception of quantitative non-viewers, had significant ($p < 0.05$) improvement given the Part 1/Part 2 paired *T*-tests. *T*-tests showed significant improvement in male quantitative viewers and male non-quantitative

non-viewers. The Tukey procedure showed female quantitative viewers recorded the highest degree of 5B+ change relative to male quantitative non-viewers and female non-quantitative non-viewers.

Female subjects, regardless of functional area, were *Principle Substitution* types. Male subjects, however, did not go through the CMD stages as uniformly as female subjects. If exposed to Milgram, quantitative males increased their 5B+ scores. If not exposed, quantitative males had no change in 5A, 5B+ or *P*-scores. Non-quantitative males suffered a decline in 5A scores if exposed to Milgram. Non-quantitative males who did not view Milgram increased their use of Stage 5B+. Thus, quantitative male viewers and non-quantitative male non-viewers were *Principle Addition* types. Quantitative male non-viewers were *Retrenchment* types. Non-quantitative male viewers were *Rule-Response* types. In other words, response among males was very much contingent on their functional area of study.

Finally, the Part 2 *P*-score, differences among the viewing groups (an *F*-score of 4.63) indicates a significant difference between groups. The Tukey Procedure for viewers *P*-scores shows that female quantitative viewers scored the highest relative to both groups of male viewers. The only viewer group that scored a significant change in the paired *T*-tests was female quantitative viewers.

Hypotheses results

There were positive results from viewing Milgram. We looked for those groups where there were different responses to viewing Milgram. The results are as follows:

H1: Viewing Milgram had a positive impact on subjects' cognitive moral development. Based on the significant difference in changes of viewers versus non-viewers 5A and 5B+ scores, we must reject the null hypothesis that there was no effect. Even though all subjects employed *Principle Substitution*, the change in viewers was significantly greater.

H2: There is a significant difference in the impact Milgram had on female versus male subjects' CMD. Based on significant differences in Part 2, 5A and *P*-scores, as well as results

indicating females 5A and 5B+ scores exhibited the highest degree of change, we must reject the null hypothesis that there is no difference between males and females.

*H3:*_N There is no significant difference in the impact Milgram had on quantitative-oriented subjects' CMD versus non-quantitative-oriented subjects' moral development. Since there were no statistically significant differences in Part 2 scores or change scores observed in quantitative versus non-quantitative oriented viewers, we must accept the null hypothesis.

H4: There is a significant difference in the impact Milgram had on subjects' CMD in different functional areas by gender. Based on a significant difference in Part 2 viewer *P*-scores we must reject the null hypothesis. Among viewers, quantitative females Part 2 *P*-scores were significantly higher than for both groups of males. While female viewers consistently used *Principle Substitution*, quantitative males used *Principle Addition* and non-quantitative males used *Rule Response*

Discussion and conclusion

Viewing Milgram had an impact on subjects and affected each gender differently. Results for eight groups demonstrate that: (1) females followed consistent (*Principle Substitution*) patterns in developing moral reasoning (males followed different and varied patterns) and; (2) the end result of tests among quantitative female viewers were higher 5B+ and *P*-scores relative to other viewer groups.

Subjects in this study were not unusual – they had an average *P*-score of 40.5 – lower than the average college student of 42.3, but not a significant difference. A finer grained analysis, starting with gender differences, yields a more interesting picture. Consistent with other work (e.g. Gilligan, 1977, 1982) this study found that men and women differed on how they resolved moral conflicts – e.g. women viewers outscored men in Part 2 *P*-score. Why were these women's scores different? To answer this question we looked at another dimension of the

sample: quantitative and non-quantitative areas of study.

We found that differences in area of study alone had a very minor impact. This may explain why Stanga and Turpen's (1991) study of accounting professionals and Derry's (1987, 1989) studies of managers found no gender differences in ethical values. Looked at within their functional areas, men and women may show little difference. Yet looking at the differences between functional orientations along with gender did highlight differences – e.g. females consistent use of what we called a *Principle Substitution* pattern of moral growth versus males range of patterns that varied by functional area of study.

Given the *T*-test results between Parts 1 and 2, *P*-score alone did not show any significant change in subjects' moral growth brought about via their exposure to Milgram. This is consistent with Rest's (1994) assertion that short courses or interventions have little impact on cognitive moral development (at least as measured by the *P*-score). Unlike Rest's conclusions, the results of this study demonstrate that significant learning is still occurring. The underlying *P*-score components showed significant shifts (given the *T*-test results between Parts 1 and 2); these underlying components of the *P*-score also showed significant differences in the manner in which people developed various approaches to moral problems.

In this study, when we confronted subjects with new information having a significant moral dimension (the viewing of the Milgram video), they responded with a change in Stage 5A scores (a statistically significant decrease) and 5B+ scores (a statistically significant increase) than did the overall subject pool. When we segmented subjects by gender, it was only the male non-viewer group that did not show any significant change in Stage 5A or 5B+. Likewise, when the groups were segmented into quantitative/non-quantitative classifications, it was only the quantitative non-viewer group that showed a lack of significant response across all three measures (5A, 5B+ and P).

Although Lord and DeZoort (2001) found that obedience pressure among auditors created a willingness to sign-off on materially misstated account balances, this study found that viewing Milgram did impact subjects' moral judgements when confronted with information regarding obedience pressures. When all eight groups were considered, it was the

quantitative *non-viewer* groups (both male and female), which did not show any significant positive change given the *T*-tests over all three measures.

What does this mean for those who may end-up needing to make sensitive moral decisions with the potential to bring down a company? If Wimalasiri et al. (1996), and Wimalasiri (2001) are correct about the rough equivalence in CMD scores between students and managers, one conclusion may be to include diverse groups of men and women with both quantitative and non-quantitative backgrounds in decision-making processes. This is consistent with the Hambrick et al. (1996) call for heterogeneous top management teams. Thus, diverse backgrounds may be useful in overseeing decisions involving the potential to put organizations in critically risky situations. Given that different functional orientations may result in different CMD growth paths, a range of perspectives may be useful in aiding in the development of better moral judgements to critical organizational challenges.

In spite of possible worries over group decision failures, e.g. groupthink (Janis, 1983), group decisions have been found to aid in moral reasoning. Nichols and Day (1982) found that when groups composed of business students who scored at a higher level of CMD were brought together they could arrive at more postconventional arguments to support their decisions.

Future directions

There are some important points to note here in terms of future research directions. First, given what was found in the current study, and consistent with Nelson's (2000) call to do so, future research needs to take a finer grained view of cognitive moral development. Researchers need to look at stage and substage components of DIT scores as a more detailed measure of a person's stage of CMD.

Second, there is a need to look at the impact of short run ethical education interventions (particularly given a person's stage and substage components). In general, there is a need to look at how people develop through both A and B substages given various treatments; when substages are studied – as was done in the current study – significant learning was found to occur. However, longer-term

longitudinal studies may need to be undertaken to see if the results found here endure for significant periods of time.

Third, when looking at CMD one needs to look beyond simple professional designation. Within any profession there may be sub-disciplines or orientations that may be more important in terms of their impact on CMD than any overall professional designation. This is what we discovered when our sample was stratified into quantitative/non-quantitative groups. Use of dilemmas oriented to different functional areas may illuminate a greater degree of difference between groups.

Finally, some subjects in this study had already taken a number of business classes. To what extent

the CMD of these subjects was impacted by their programme of study as opposed to having a pre-disposition to developing in the way they did would also be an area for further investigation. In other words, “Is it the business-people we get, or what we teach them, that makes them more moral or not?”

Acknowledgements

We would like to acknowledge the assistance of Jack Dunning of Algoma University, Tom Jones of the University of Washington, Chris Hamblin of Simon Fraser University and the two anonymous reviewers.

Appendix 1. Breakdown for viewers/non-viewers, by gender & functional area of study

		Part 1 Score			Part 2 Score		T	Change			
		n	Mean	(s)	Mean	(s)	Score	Mean	(s)		
Viewers vs. Non Viewers	5A Score	Viewed	39	40.90	(21.13)	19.90	(10.72)	-5.90**	-21.00	(22.23)	
		Unviewed	35	36.11	(20.30)	26.31	(16.18)	-2.35*	-9.80	(24.69)	
		T Score	—	0.99	—	-2.01*	—	—	-2.04*	—	
	5B+ Score	Viewed	39	13.10	(16.54)	41.46	(20.52)	7.29**	28.36	(24.28)	
		Unviewed	35	20.62	(18.78)	34.57	(18.97)	3.50**	13.94	(23.74)	
		T Score	—	-1.82	—	1.50	—	—	-2.59**	—	
	P-Score	Viewed	39	37.72	(15.43)	40.00	(14.05)	0.71	2.28	(20.06)	
		Unviewed	35	38.14	(17.17)	40.46	(16.76)	0.67	2.31	(20.31)	
		T Score	—	-0.11	—	-0.13	—	—	-0.01	—	
Group Break-down by Gender	5A Score	Male	Viewed	12	35.08	(16.89)	14.17^L	(10.35)	-3.60**	-20.91	(20.11)
			Unviewed	21	31.71	(20.86)	28.14^H	(16.92)	-0.76	-3.57^L	(21.54)
		Female	Viewed	27	43.48	(22.56)	22.44	(10.04)	-4.66**	-21.04^H	(23.48)
			Unviewed	14	42.71	(18.15)	23.57	(15.18)	-2.66*	-19.14	(26.91)
	F Score (Viewer's T)	—	1.61	(-1.29)	2.80*	(2.33)*	—	2.71*	(-0.02)		
		5B+ Score	Male	Viewed	12	13.58	(12.59)	32.08	(21.60)	2.68*	18.50
	Unviewed			21	16.28	(20.06)	31.48	(18.72)	2.62*	15.19	(26.58)
	Female		Viewed	27	12.89	(18.23)	45.63	(18.96)	7.22**	32.74^H	(23.55)
			Unviewed	14	27.14	(15.09)	39.21	(19.05)	2.39*	12.07^L	(18.92)
	F Score (Viewer's T)	—	2.22	(0.14)	2.58	(-1.88)	—	3.31*	(-1.72)		
		P-Score	Male	Viewed	12	33.67	(12.79)	30.00^L	(14.17)	-0.59	-3.67
	Unviewed			21	32.52	(17.96)	39.81	(16.63)	1.79	7.29	(18.65)
Female	Viewed		27	39.52	(16.37)	44.44^H	(11.70)	1.34	4.92	(19.16)	
	Unviewed		14	46.57	(12.12)	41.43	(17.54)	-0.91	-5.14	(21.07)	
F Score (Viewer's T)	—	2.65	(-1.21)	2.68*	(3.09)*	—	1.63	(1.19)			
	Break-down Functional area of study	5A Score	Quant.	Viewed	22	39.14	(22.78)	20.41	(9.05)	-4.07**	-18.73
Unviewed				17	30.41	(19.10)	22.71	(12.87)	-1.43	-7.71	(22.25)
Non			Viewed	17	43.18	(19.22)	19.24	(12.83)	-4.22**	-23.94	(23.37)
			Unviewed	18	41.50	(20.43)	29.72	(18.50)	-1.83	-11.78	(27.29)
F Score (Viewer's T)		—	1.30	(-0.60)	2.19	(0.32)	—	1.62	(-0.71)		
		5B+ Score	Quant.	Viewed	22	11.77	(13.37)	44.91	(18.65)	7.41**	33.14^H
Unviewed				17	22.06	(21.28)	33.65	(20.31)	1.66	11.59^L	(28.82)
Non			Viewed	17	14.82	(20.23)	37.00	(22.92)	3.34**	22.18	(27.40)
			Unviewed	18	19.28	(15.31)	35.44	(18.16)	3.85**	16.16	(17.80)
F Score (Viewer's T)		—	1.26	(-0.54)	1.27	(1.17)	—	3.02*	(1.37)		
		P-Score	Quant.	Viewed	22	35.55	(16.20)	42.45	(13.98)	1.76	6.90
Unviewed				17	34.53	(19.02)	37.18	(16.60)	0.57	2.65	(19.15)
Non	Viewed		17	40.53	(14.37)	36.82	(13.99)	-0.73	-3.71	(20.99)	
	Unviewed		18	41.56	(14.95)	43.56	(16.79)	0.39	2.00	(21.90)	
F Score (Viewer's T)	—	0.85	(-1.02)	0.94	(1.25)	—	0.89	(1.65)			

* p < .05 ** p < .01 Tukey Procedure: ^H Highest group ^L Lowest group

Appendix 2. Breakdown for viewers/non-viewers for all 8 groups

			n	Part 1 Score		Part 2 Score		T Score	Change		
				Mean	(s)	Mean	(s)		Mean	(s)	
5 A Score	Male	Quant.	Viewed	6	31.67	(20.66)	16.17	(9.83)	-1.95	-15.50	(19.45)
			Unviewed	11	26.45	(20.70)	21.10	(9.82)	-0.86	-5.36	(20.75)
		Non-Quant.	Viewed	6	38.50	(13.16)	12.17^L	(11.37)	-3.07*	-26.33	(21.00)
			Unviewed	10	37.50	(20.49)	35.90^H	(20.04)	-0.22	-1.60	(23.33)
	Female	Quant.	Viewed	16	41.94	(23.52)	22.00	(8.52)	-3.50*	-19.94	(22.80)
			Unviewed	6	37.67	(14.56)	25.67	(17.92)	-1.12	-12.00	(26.26)
		Non-Quant.	Viewed	11	45.73	(22.00)	23.09	(12.35)	-2.95*	-22.64	(25.46)
			Unviewed	8	46.50	(20.56)	22.00	(13.87)	-2.49*	-24.50	(27.85)
	F Score	Viewers	39	—	0.59	—	1.95	—	—	0.25	—
	F Score	All	74	—	1.06	—	2.30*	—	—	1.39	—
5B+ Score	Male	Quant.	Viewed	6	13.50	(14.72)	28.00	(9.78)	2.96*	14.50	(11.98)
			Unviewed	11	17.91	(25.73)	26.45^L	(18.56)	0.91	8.55^L	(30.98)
		Non-Quant.	Viewed	6	13.67	(11.48)	36.17	(29.85)	1.68	22.50	(32.83)
			Unviewed	10	14.50	(12.30)	37.00	(18.22)	3.60**	22.50	(32.83)
	Female	Quant.	Viewed	16	11.13	(13.28)	51.25^H	(17.26)	8.27**	40.13^H	(19.41)
			Unviewed	6	29.67	(12.54)	46.83	(17.57)	1.61	17.16	(26.10)
		Non-Quant.	Viewed	11	15.45	(24.25)	37.45	(19.06)	2.91*	22.00	(25.74)
			Unviewed	8	25.25	(17.35)	33.50	(19.15)	1.98	8.25^L	(11.80)
	F Score	Viewers	39	—	0.14	—	2.65	—	—	2.55	—
	F Score	All	74	—	1.02	—	2.21*	—	—	2.43*	—
P- Score	Male	Quant.	Viewed	6	31.00	(10.79)	28.83^{VL}	(10.34)	0.37	-2.17	(14.36)
			Unviewed	11	29.36	(21.31)	31.55	(14.14)	0.37	2.18	(19.69)
		Non-Quant.	Viewed	6	36.33	(15.06)	31.17^{VL}	(18.21)	-0.44	-5.16	(28.54)
			Unviewed	10	36.00	(13.66)	48.90	(14.75)	2.46*	12.90	(16.59)
	Female	Quant.	Viewed	16	37.25	(17.81)	47.56^{VH}	(11.66)	2.17*	10.31	(19.05)
			Unviewed	6	44.00	(9.10)	47.50	(16.85)	0.43	3.50	(19.92)
		Non-Quant.	Viewed	11	42.82	(14.16)	39.91	(10.65)	-0.56	-2.91	(17.20)
			Unviewed	8	48.50	(14.27)	36.88	(17.71)	1.59	-11.62	(20.72)
	F Score	Viewers	39	—	0.79	—	4.63**	—	—	1.53	—
	F Score	All	74	—	1.44	—	3.00**	—	—	1.69	—

* p < .05 ** p < .01 Tukey Procedure All groups: ^H Highest group ^L Lowest group
 Viewers only: ^{VH} Highest group ^{VL} Lowest group

Notes

¹ Throughout this study we will use the term “gender” to describe the subjects’ sex. While gender is a socio-psychological concept and sex refers to whether someone is male or female, gender is the more common term and studies in this area may even use both terms to refer to a person’s sex (e.g. see Bebeau, 2002).

² Such an indirect route would explain apparent retrograde moves in some individual’s moral development scores and allow for a model that moves an individual steadily forward through the moral development stages.

³ Guthrie (2001) found a correlation between field independence, locus of control and CMD. In Guthrie’s study, externally controlled, field dependent students scored significantly lower in their stage of CMD than internally controlled, field independent subjects.

⁴ Inferior moral judgements may stem from the nature of the *judgements* or the *people* making them. Business *judgements* may be addressed at lower moral development stages. Weber (1990a) and Carpendale and Krebs (1992) found the moral reasoning stage used was lower in business contexts. *People* choosing to go into management may have lower moral development levels (Hiltebeitel and Jones, 1991, 1992; Ponemon and Gabhart, 1994). Managers may also view business as a game where amoral actions are accepted (Reall et al., 1998), or techniques learned in business schools (e.g. cost/benefit analysis) create psychological distance between managers and others leading to amoral behaviour (Nicholson, 1998). Also, more moral individuals may select themselves out of management (Jones and Quinn, 1995) or opt out due to workplace stress (Mason and Mudrack, 1997).

⁵ Regarding job characteristics as a moderator of moral behaviour, Akaah (1996) found marketing executives showed a higher frequency of ethical judgements than those performing a marketing research role. Lincoln et al. (1982) used 11 cases to assess the ethical beliefs and values of Fortune 500 executives and found significant differences between those in different functional areas. Corporate marketing vice presidents felt their ethics were less compromised by organizational pressures than production vice presidents (VPs). In turn, production VPs felt their ethics were less compromised by such pressures than finance VPs.

⁶ Schemas are sets of cognitive constructs that depict "...organized knowledge about a given concept or type of stimulus" (Fiske and Taylor, 1984: 140).

⁷ While exposure to Milgram may aid in creating moral development schema that makes future managers more sensitive to recognizing a moral issue, our main concern here is whether Milgram affects their moral judgement.

⁸ Rest et al. (1999), argue for use of the label "post-conventional" to replace "principled" moral reasoning, and we have done so here.

⁹ In a review DIT studies, King and Mayhew (2002) discuss 13 studies with interventions intended to change students' moral development scores. In 11 of 13 studies, researchers, using a wide range of methods were able to show significant improvements in students' scores. King and Mayhew make special note of Boss's (1994) study because it, like the current study, control for class size, instructor and text used across two sections of an ethics class. Boss found curriculum and discussion of moral dilemmas was effective only for the class where students were completed 20 hours of community service work directly with those in need and kept a journal.

¹⁰ Each gender gives some 'care' and some 'justice' responses. Reed (1997: 227) notes that in Gilligan's initial studies, half the subjects gave responses that were at least 75% 'justice' based; 34% of subjects gave responses at least 25% 'justice' based; only 16% of (12 women and one man) gave 'care' responses more than 75% of the time. So even in Gilligan's work, 'morality' is weighted towards a justice perspective. Use of the 'care' perspective is more reflective of the type or setting of the moral dilemma than gender. "Dilemmas located within a 'community' or 'family' context are likely to invoke caring and response considerations... In brief, choice of orientation seems to be primarily a function of setting and dilemma, not sex" (Kohlberg et al., 1983 – in Kohlberg, 1984: 350).

¹¹ Jones and Gaultschi (1988) found women and men do not generally show much difference in ethical attitudes. Kidwell et al. (1987) found no significant differ-

ences between the ethical perceptions of men and women. Sikula and Costa (1994) found no differences among male and female college students' ethical values. Stanga and Turpen's (1991) study of accounting professionals found no gender differences in ethical values.

¹² Weber and Wasieleski (2001) account for this difference as an organizational influence. Organizations would play a role but such differences may also be found in functional area orientations. Thus, the roots and route to such variation may be found in the way that people with different functional area orientations develop morally.

¹³ Intervention studies with DIT comparisons between majors within business could not be located. Placing the sample's two economics majors in the non-quantitative group did not change the hypotheses test results.

¹⁴ These include most prominently, Kohlberg's Moral Judgement Interview or MJI (see Colby and Kohlberg, 1987), an abbreviated version of the MJI by Weber (see Weber, 1990a), Gibbs' Social Reflection Questionnaire (see Gibbs et al., 1982) and Rest's Defining Issues Test or DIT (see Rest, 1979), among others.

¹⁵ One drawback of the DIT is its lack of business dilemmas (as called for by Weber, 1992). Unless very carefully written, business dilemmas may impact different functional areas differently, where a standardized test would be less open to question. Also the moral implications of obedience to authority are of interest across a wide range of organizations and use of business dilemmas may limit the ability to reproduce this study's results and generalize its findings to other fields.

¹⁶ Three DIT scenarios have four statements that address Kohlberg's postconventional morality stages and each of these three scenarios have a maximum of ten points (4 + 3 + 2 + 1). For each postconventional choice points were awarded in accordance with Rest's (1994) manual and system: four points for a first choice, three for a second choice, two points for third choice and one point for fourth choice. Three scenarios have three statements that address the same stages and thus each of these three scenarios have a maximum of nine points (4 + 3 + 2).

¹⁷ Part 1 DIT scenarios were Heinz and the drug, Student take-over and Escaped prisoner. Part 2 scenarios were Doctor's dilemma, Webster and Newspaper. See Rest (1986) for details.

¹⁸ Using these additional measures makes sense given Mudrack's (2003) admonition against "...sole reliance on P-scores [since it] may tend to obscure the magnitude of actual relationships with other variables of interest."

¹⁹ An Excel spreadsheet was developed where subjects' choices were recorded and Part 1 and Part 2 5A,

5B+ and *P*-scores raw scores were calculated from subjects' choice of Stage 5A, 5B, and 6 selections in accordance with Rest's (1994) manual and system that awards, four points for a first choice, three for a second choice, two points for third choice and one point for fourth choice. Thus a maximum raw score for Stage 5B+ Part 1 is 15, for Part 2 it is 18. To prevent the analysis from reflecting changes from different maximums for each part, the scores were transformed to have the standard DIT maximum of 95 for each part (Part 1 score = (Part 1 raw score/15) × 95; Part 2 score = (Part 2 raw score/18) × 95).

²⁰ Two other possibilities are: 5A and 5B+ both increase, or 5A increases and 5B+ decreases. The first is unlikely as there would be more short-term intervention studies with a positive *P*-score impact. The latter is unlikely, as it is the opposite of what is expected from viewing Milgram – less, not more, independence in moral judgment.

²¹ To insure that changes in test subjects were not the result of differences between Parts 1 and 2 of the DIT, the entire DIT was administered to a small group of pretest subjects. There were no statistically significant differences ($p < 0.05$) in scores between Parts 1 and 2 for these subjects.

²² Wimalasiri's (2001) study of 159 managers and 89 management students in Australia showed no statistical difference between the groups' DIT scores ($t = 0.09$). Wimalasiri et al. (1996) found the same lack of significant differences. *P*-scores of the students in the current study did not differ significantly from adults ($t = 0.11$) in Rest's initial 1979 sample (see Rest, 1994: 6.1). There was no significant difference between *P*-scores of Uddin and Gillett's (2002) CFO subjects and subjects in the current study majoring in quantitative areas ($t = 0.18$).

²³ Disparities in age and work experience may impact the results (Weber, 1990b). See King and Mayhew (2002) for a summary of longitudinal studies regarding age and work experience. The dropped subjects here were outliers in that their age (26–48) averaged 16 years older than subjects kept. With an average of 11 years full-time work experience (range 4–26) dropped subjects had 10 years more work experience than subjects kept (range 0–3). There was no correlation found between *P*, 5A and 5B+ scores and university or business credits earned.

²⁴ The film shows that when subjects were part of a group that wanted to go to on, obedience was greater (especially if the subject did not administer the shock). Where a researcher provided only initial commands or left the instructions in an envelope, obedience was less. When the subject was in the same room or had to hold

a victim's hand to a shock plate, obedience was lower still. In groups that wanted to stop, subject obedience was near zero.

²⁵ This does not mean that such skills will be maintained. Babies may take a few steps and be at a developmental stage where they are able to walk. However, if for some reason they are not permitted to practice walking they are unlikely to maintain that level of development.

²⁶ Results of analyses with the five dropped subjects included did not alter the results of the hypotheses tests.

²⁷ Though results for *T*-tests and ANOVAs are shown in the appendices, non-parametric tests (Wilcoxon Matched-pairs Signed-ranks tests and Kruskal–Wallis one-way ANOVAs) showed significant results on the same measures.

References

- Akaah, I.: 1996, 'The Influence of Organizational Rank and Role on Marketing Professionals' Ethical Judgments', *Journal of Business Ethics* **15**, 605–613.
- Beard, P. N.: 1983, 'The Effect of Moral Intervention on Different Student Population Groups', *Humanitas: Journal for Research in the Human Sciences* **9**(3), 355–363.
- 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.
- Behr, P. and A. Witt: 2005, 'Visionary's Dream Led to Risky Business', in A. Gini (ed.), *Case Studies in Business Ethics, 5th Edition* (Pearson Prentice/Hall, Upper Saddle River NJ), pp. 385–394.
- Beltramini, R. F., R. A. Peterson and G. Kozmetsky: 1984, 'Concerns of College Students Regarding Business Ethics', *Journal of Business Ethics* **3**, 195–200.
- Bernardi, R. A. and D. F. Arnold, Sr.: 1997, 'An Examination of Moral Development within Public Accounting by Gender, Staff level, and Firm', *Contemporary Accounting Research* **14**(4), 653–668.
- Borkowski, S. C. and Y. J. Ugras: 1998, 'Business Students and Ethics: A Meta-analysis', *Journal of Business Ethics* **17**(11), 1117–1127.
- Boss, J. A.: 1994, 'The Effect of Community Service Work on the Moral Development of College Ethics Students', *Journal of Moral Education* **23**, 183–198.
- Carpendale, J. I. and D. L. Krebs: 1992, 'Situational Variation in Moral Judgment: In a Stage or on a Stage?', *Journal of Youth and Adolescence* **21**(2), 203–225.

- Cherry, S.: 2000, Course Notes 12: Inferences for factor level means. Analysis of Variance & Experimental Design, <http://www.math.montana.edu/~cherry/st412/notes/Ho9/Ho9.html> (accessed Oct. 10, 2002).
- Colby, A. and L. Kohlberg: 1987, *The Measurement of Moral Judgment: Standard Issue Scoring Manual* (Cambridge University Press, Cambridge).
- Derry, R.: 1987, 'Moral Reasoning in Work-related Conflicts', in W. C. Frederick (eds.), *Research in Corporate Social Performance and Policy, Vol 9* (JAI, Greenwich Conn), pp. 25–50.
- Derry, R.: 1989, 'An Empirical Study of Moral Reasoning among Managers', *Journal of Business Ethics* **8**, 855–862.
- Elm, D. and J. Weber: 1994, 'Measuring Moral Judgment: The Moral Judgment Interview or the Defining Issues Test?', *Journal of Business Ethics* **13**(5), 341–355.
- Fiske, S. T. and S. E. Taylor: 1984, *Social Cognition* (Addison-Wesley, Reading, Mass).
- Gautschi, F. H. and T. M. Jones: 1998, 'Enhancing the Ability of Business Students to Recognize Ethical Issues', *Journal of Business Ethics* **16**, 205–216.
- Gibbs, J. C., K. F. Widaman and A. Colby: 1982, *Social Intelligence: Measuring the Development of Sociomoral Reflection*. (Prentice-Hall, Inc., Englewood Cliffs, N. J.).
- Gilligan, C.: 1977, 'In a Different Choice: Women's Conceptions of the Self and of Morality', *Harvard Educational Review* **47**, 481–517.
- Gilligan, C.: 1982, 'Adult Development and Women's Development: Arrangements for a Marriage', in J. Giele (eds.), *Women in the Middle Years* (John Wiley & Sons, New York), pp. 57–83.
- Gilligan, C. and J. Attanucci: 1988, 'Two Moral Orientations: Gender Differences and Similarities', *Merrill-Palmer Quarterly* **34**(3), 223–237.
- Gilligan, C., J. Ward, J. Taylor and B. Bardige: 1988, *Mapping the Moral Domain* (Harvard University Press, Cambridge, Mass).
- Greising, D.: 1989, Chicago's B-School Goes Touchy-Feely. *Business Week*, Nov. 27, 140.
- Guthrie, K. H.: 2001, 'Locus of Control and Field Independence-Dependence as Factors in the Development of Moral Judgment', *Journal of Genetic Psychology* **146**(1), 13–18.
- Hambrick, D. C., T. S. Cho and M. J. Chen: 1996, 'The Influence of Top Management Team Heterogeneity on Firms' Competitive Moves', *Administrative Science Quarterly* **41**, 659–684.
- Hiltebeitel, K. M. and S. K. Jones: 1991, 'Initial Evidence on the Impact of Integrating Ethics into Accounting Education', *Issues in Accounting Education* **6**(2), 262–275.
- Hiltebeitel, K. M. and S. K. Jones: 1992, 'An Assessment of Ethics Instruction in Accounting Education', *Journal of Business Ethics* **11**, 37–46.
- Hobel, J.: 2002, 'The Two Sides of HR', *Canadian HR Reporter* **15**(2), 4–8.
- Holstein, C. B.: 1976, 'Irreversible, Stepwise Sequence in the Development of Moral Judgment: A Longitudinal Study of Males and Females', *Child Development* **47**, 51–61.
- Janis, I. L.: 1983, *Groupthink: Psychological Studies of Policy Decisions and Fiascoes 2* (Houghton Mifflin, Boston).
- Jeffrey, C.: 1993, 'Ethical Development of Accounting Students, Non-accounting Business Students, and Liberal Arts Students', *Issues in Accounting Education* **8**(1), 86–98.
- Jones, T. M.: 1991, 'Ethical Decision Making by Individuals in Organizations: An Issue-Contingent Model', *Academy of Management Review* **16**, 366–395.
- Jones, T. and F. Gautschi: 1988, 'Will the Ethics of Business Change?: A Survey of Future Executives', *Journal of Business Ethics* **7**, 231–248.
- Jones, T. M. and D. P. Quinn: 1995, 'An Agent Morality View of Business Policy', *Academy of Management Review* **20**, 22–42.
- Kennedy, E. J. and L. Lawton: 1998, 'Religiousness and Business Ethics', *Journal of Business Ethics*, **17**(2): 163–175.
- Kidwell, J., R. I. Stevens and A. Bethke: 1987, 'Differences in Ethical Perceptions between Male and Female Managers: Myth or Reality?', *Journal of Business Ethics* **6**, 489–93.
- King, P. M. and M. J. Mayhew: 2002, 'Moral Judgement Development in Higher Education: Insights from the Defining Issues Test', *Journal of Moral Education* **31**(3), 247–271.
- Kohlberg, L.: 1969, 'Stage and Sequence: The Cognitive-Developmental Approach to Socialization', in D. Goslin (eds.), *Handbook of Socialization Theory and Research*. (Rand McNalley, Chicago), pp. 347–480.
- Kohlberg, L.: 1981, *The Philosophy of Moral Development, Essays in Moral Development* (Harper & Row, New York).
- Kohlberg, L.: 1984, *Essays in Moral Development, the Psychology of Moral Development, Vol II* (Harper & Row, New York)..
- Kohlberg, L., C. Levine and A. Hewer: 1983, Moral Stages: A Current Formulation and a Response to Critics. *Contributions to Human Development*, 10. Basel: S Karger. Reprinted in Kohlberg, 1984.
- Lincoln, D. J., M. M. Pressley and T. Little: 1982, 'Ethical Beliefs and Personal Values of Top Level Executives', *Journal of Business Research* **10**, 475–487.

- Lord, A. T. and F. T. DeZoort: 2001, 'The Impact of Commitment and Moral Reasoning on Auditors' Responses to Social Influence Pressure', *Accounting, Organizations and Society* **26**(3), 215–235.
- Mason, E. S. and P. E. Mudrack: 1997, 'Do Complex Moral Reasoners Experience Greater Ethical Work Conflict?', *Journal of Business Ethics* **16**, 1311–18.
- Miceli, N. S.: 1996, 'Deviant Managerial Behavior: Costs, Outcomes and Prevention', *Journal of Business Ethics* **15**(6), 703–710.
- Milgram, S.: 1974, *Obedience to Authority* (Harper & Row, New York).
- Milgram, S.: 1965, *Obedience to Authority* (New York University Film Library, New York).
- Moberg, D. and A. E. Romar: 2005, 'The Good Old Boys at WorldCom', in A. Gini (eds.), *Case Studies in Business Ethics, 5th Edition* (Pearson Prentice/Hall, Upper Saddle River NJ), pp. 205–213.
- Mudrack, P. E.: 2003, The Untapped Relevance of Moral Development Theory in the Study of Business Ethics. *Journal of Business Ethics*, 42(3): 225–238.
- Nelson, D.: 2000, 'Multiple Methods for Analyzing Moral Judgment Development Using the Defining Issues Test', in Ron Burwell (eds.), *Research on Christian Higher Education, Vol. 7* (Council for Christian Colleges & Universities, Washington, D.C.).
- Neter, J., M. H. Kutner, W. Wasserman and C. J. Nachtsheim: 1996, *Applied Statistical Models, Fourth Edition* (McGraw-Hill, New York)..
- Nicholson, N.: 1998, 'How Hardwired is Human Behavior', *Harvard Business Review* **76**(4), 134–147.
- Nichols, M. L. and V. E. Day: 1982, 'A Comparison of Moral Reasoning of Groups and Individuals on the 'Defining Issues Test'', *Academy of Management Journal* **25**, 201–208.
- Nisbett, R. E. and L. D. Ross: 1980, *Human Inference: Strategies and Shortcomings of Social Judgment* (Prentice-Hall, Englewood Cliffs NJ).
- Nisan, M.: 1984, 'Social Norms and Moral Judgment', in W. Kurtines and J. Gewirtz (eds.), *Morality and Moral Development* (Wiley, New York), pp. 208–224.
- Nocera, J.: 2002, 'System Failure, Corporate America has Lost Its Way. Here's a Road Map for Restoring Confidence', *Fortune* **145**(13), 62–74.
- Nucci, L. P.: 1982, 'Children's Responses to Moral and Social Conceptual Transgressions in Free-Play Settings', *Child Development* **53**, 1337–1342.
- Peterson, R. A., R. F. Beltramini and G. Kozmetsky: 1991, 'Concerns of College Students Regarding Business Ethics: A Replication', *Journal of Business Ethics* **10**, 733–739.
- Ponemon, L. A. and D. Gabhart: 1994, 'Ethical Reasoning Research in the Accounting and Auditing Professions', in J. R. Rest and D. Narvaez (eds.), *Moral Development in the Professions: Psychology and Applied Ethics* (Lawrence Erlbaum Associates, Hillsdale, N.J.), pp. 101–119.
- Pritchard, R. E., G. C. Potter and M. Saccucci: 2004, 'The Selection of a Business Major: Elements Influencing Student Choice', *Journal of Education for Business* **79**(3), 152–156.
- Reall, M. J., J. J. Bailey and S. K. Stoll: 1998, 'Moral Reasoning on Hold during a Competitive Game', *Journal of Business Ethics* **17**, 1205–1210.
- Reed, D. R. C.: 1997, *Following Kohlberg: Liberalism and the Practice of Democratic Community* (University of Notre Dame Press, Notre Dame, Indiana).
- Rest, J. R.: 1979, *Development in Judging Moral Issues* (University of Minnesota Press, Minneapolis).
- Rest, J. R.: 1986, *Moral Development: Advances in Research and Theory* (Praeger, New York).
- Rest, J. R.: 1994, *Revised Manual for the Defining Issues Test* (Univ of Minn. Press, Minneapolis)..
- Rest, J., D. Narvaez, M. J. Bebeau and S. J. Thoma: 1999, *Postconventional Moral Thinking: A Neo-Kohlbergian Approach* (Lawrence Erlbaum Associates, Mahwah, NJ)..
- Rifkin, G.: 1996, 'One Man's Search for Leadership', *Forbes* **April, 8**, 110–115.
- Schlaefli, A., J. Rest and S. Thoma: 1985, 'Does Moral Education Improve Moral Judgment? A Meta-Analysis of Intervention Studies Using the Defining Issues Test', *Review of Educational Research* **55**, 319–352.
- Self, D. J., D. C. Baldwin, Jr. and M. Olivarez: 1993, 'Teaching Medical Ethics to First-year Students by Using Film Discussion to Develop their Moral Reasoning', *Academic Medicine* **68**, 383–385.
- Sikula, A. and A. Costa: 1994, 'Are Women More Ethical Than Men?', *Journal of Business Ethics* **13**, 859–871.
- Smith, H. R. and A. B. Carroll: 1984, 'Organizational Ethics: A Stacked Deck', *Journal of Business Ethics* **3**, 95–100.
- Sneed, J. and D. A. Morgan: 1999, 'Evaluating the Verbal, Quantitative, and Problem-Solving Skills of Students Entering the Accounting Curriculum', *Management Research News* **22**(4), 22–28.
- St. Pierre, K., E. Nelson and A. Gabbini: 1990, 'A Study of the Ethical Development of Accounting Majors in Relation to other Business and Nonbusiness Disciplines', *Accounting Educators Journal* **8**, 23–25.
- Stanga, K. G. and R. C. A. Turpen: 1991, 'Ethical Judgments on Selected Accounting Issues: An Empirical Study', *Journal of Business Ethics* **10**, 739–747.
- 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**(4), 659–669.

- Trevino, L. K.: 1986, 'Ethical Decision Making in Organizations: A Person-Situation Interactionist Model', *The Academy of Management Review* **11**(3), 601-617.
- Uddin, N. and P. R. Gillett: 2002, 'The Effects of Moral Reasoning and Self-monitoring on CFO Intentions to Report Fraudulently on Financial Statements', *Journal of Business Ethics* **40**(1), 15-21.
- Walker, L.: 1984, 'Sex Differences in the Development of Moral Reasoning: A Critical Review', *Child Development* **55**, 677-691.
- Weber, J.: 1990a, 'Managers' Moral Reasoning: Assessing Their Responses to Three Moral Dilemmas', *Human Relations* **43**(7), 687-702.
- Weber, J.: 1990b, 'Measuring the Impact of Teaching Ethics to Future Managers: A Review, Assessment and Recommendations', *Journal of Business Ethics* **9**(3), 183-190.
- Weber, J.: 1991, 'Adapting Kohlberg to Enhance the Assessment of Managers' Moral Reasoning', *Business Ethics Quarterly* **1**(3), 293-318.
- Weber, J.: 1992, 'Scenarios in Business Ethics Research: Review, Critical Assessment, and Recommendations', *Business Ethics Quarterly* **2**(2), 137-160.
- Weber, J. and D. Wasieleski: 2001, 'Investigating Influences on Managers' Moral Reasoning', *Business and Society* **40**(1), 79-111.
- White, R. D.: 2002, 'Do Employees Act like They Think? Exploring the Dichotomy Between Moral Judgment and Ethical Behavior', *Public Administration Quarterly* **25**(4), 391-412.
- Wimalasiri, J. S.: 2001, 'Moral Reasoning Capacity of Management Students and Practitioners: An Empirical Study in Australia', *Journal of Managerial Psychology* **16**, 614-634.
- Wimalasiri, J. S., F. Pavri and A. K. Abdul: 1996, 'An Empirical Study of Moral Reasoning Among Managers in Singapore', *Journal of Business Ethics* **15**(12), 1331-1341.

Jerry Paul Sheppard

Faculty of Business Administration,

Simon Fraser University,

8888 University Drive, Burnaby, BC,

Canada, V5A 1S6,

E-mail: sheppard@sfu.ca

Marnie Young

Adler School of Professional Psychology,

595 Burrard Street,

Vancouver BC,

Canada, Ste. 753

Copyright of *Journal of Business Ethics* is the property of Springer Science & Business Media B.V. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

Copyright of *Journal of Business Ethics* is the property of Springer Science & Business Media B.V. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.