A CALCULATED MORALITY: DEVELOPMENT AND VALIDATION OF A BUSINESS SPECIFIC TEST OF MORAL REASONING

Nathan L. Steele
Leonard Branson
University of Illinois—Springfield

ABSTRACT

Humans separate their morality or its application by situational constraints (e.g., bluffing is perfectly moral in games of poker, but not in courts of law), but it is not known if moral reasoning development holds the same weight or paths in business as other decision making environments. The present study aimed to develop a business specific test of moral reasoning called the Business Moral Reasoning Test (B-MRT) and cross-validate it against respected measures of moral reasoning development. The B-MRT possesses high internal reliability and consistency, strong correlation to the constructs of Lind’s (2014) Moral Judgment Test (MJT), as well as unique sensitivity to noted predictors of moral reasoning development such as education. Limitations and future directions as well as implications for education and business are discussed.

Keywords: Moral reasoning, business ethics, ethics education

INTRODUCTION

Preparedness to deal with situations ethically is in large part based on moral reasoning development (Rest & Navarez, 1994). General measures of moral reasoning such as the Defining Issues Test 2 (Rest, 1990) and the Moral Judgment Test (Lind, 2014) do not assess moral reasoning in the face of dilemmas like those that resulted in recent economic catastrophes, specifically business decision making.

World economies are still recovering from and often falling back into morasses created by unethical behavior committed to by many, but often by those at the very top echelons. Many of these decision makers have attained elite credentials including degrees from prominent business schools. What this seems to suggest is that thus far, the education and training offered in these programs has failed these individuals in terms of training them to deal with the decisions and environments they encounter ethically. Corporate ethics has been questioned even before these most recent economic collapses.

Companies such as Enron and WorldCom have presented cases rife with the larger negative results of immoral actions taken by a few executives for the societies and economies in which they operate. Beyond the actual economic damages, perception of business ethics in practice and in the educational arena has been called into question. The individuals largely responsible for these crises of conscience have often been in possession of elite credentials from top rated business schools.

For example, Jeffrey Skilling, former CEO of Enron, was granted an MBA from Harvard Business School (Schwartz, 2002). Andrew Fastow, the former CFO of Enron, matriculated from the prestigious Kellogg School of Management (Barboza & Schwartz, 2002). Clifford Baxter, also of Enron infamy, received his MBA from New York University, alma mater of Alan Greenspan and many other power brokers of economy and business (Jennings, 2004). Scott
Sullivan, who was named CFO of the Year by CFO Magazine in 1998 for his work at WorldCom that led to its eventual demise, graduated from the MBA program at Oswego State (Yang & Grow, 2005).

Public outcry has followed the economic gambles and unethical activities of these elite leaders of industry, and this outcry has formed a spearhead calling for greater focus on ethical education in business curricula (Desplaces, Melcher, Beauvais & Bosco, 2007; Lopez, Rechner, & Olson-Buchanan, 2005; Pettijohn, Pettijohn, & Taylor, 2008). The Association for the Advancement of Collegiate Schools of Business (AACSB), the largest accrediting body for business programs, has required the schools they accredit to develop ethical codes for their students, administrators, and faculty (AACSB, 2013, p. 5), and that business schools globally teach ethics and ethical reasoning as fundamental knowledge for students of business and management (pp. 30-33).

One may divine from these moves that the public and educators broadly suspect that more action needs to be taken to ensure that students of business in particular are the beneficiaries of deeper education in ethical practices and philosophies. The implicit assumption embodied in this push is that education, and in particular business education, can advance the ethical functioning of individuals.

As a subdivision of philosophy, ethics includes “the study of moral judgments – value judgments about what is virtuous or base, just or unjust, morally right or wrong, morally good or bad or evil, morally proper or improper” (Moore & Bruder, 2002, p. 210). More particularly, business ethics is held to be “a specialized study of right and wrong…[which]… concentrates on how moral standards apply particularly to business policies, institutions, and behavior” (Velasquez, 2006, p. 16). Schwarts and Weber (2006) specify business ethics to include any activities involving “entities operating within or related to a business context that involves the explicit interaction and/or application of ethical (i.e., moral) standards” (p. 386). These definitions imply that the gains via instruction in business ethics would be to enhance the functioning of one’s moral reasoning.

This study aimed to develop a business specific test of moral reasoning called the Business Moral Reasoning Test (B-MRT) and cross-validate it against respected measures of moral reasoning development.

**LITERATURE REVIEW**

*Moral Reasoning*

Piaget (1932) began the modern study of moral reasoning within psychology as a developmental issue, one that is not inherited, instead to be cultivated by society through education and the enforcement of norms. Developmental theories of moral reasoning long dominated the field of empirical research into its foundations and dynamics and culminated in the works of Kohlberg. Kohlberg’s (1971) work separated the development of moral reasoning into six basic stages in three broad levels—pre-conventional, conventional, and post-conventional. This theory further suggested and evidenced in empirical findings that humans progress through these stages of development linearly and progressively, never regressing to prior stages, but with the possibility that one might cease to develop further at any given stage.

The first level of development, pre-conventional, forms around the early childhood experiences of an individual and generally ends in elementary school education (Dawson, 2003).
Kohlberg proposed that in this stage of development, the individual is primarily reactive and forms an effective morality on the basis of consequences of their actions, much like the learning that takes place in classic operant conditioning, wherein punishment of moral transgressions and reward of acceptable moral behavior push and pull the individual to function or pay the price. The conventional level of development, Kohlberg’s second level, often transpires throughout adult life for the individual. In the early stages of this level, the individual recognizes what is moral by pairing abstract standards with the behaviors of others around them, and follows such standards through simple conformity pressures, but out of no higher ideal (Dawson, 2003).

In the last stage of this level of development, however, the individual begins to examine societal implications of activity and morality, and guides their activities by this larger morality toward the preservation of that societal structure and function (Dawson, 2003). The last level of development, post-conventional, is achieved by very few (Dawson). In this level, moral reasoning leads the individual to conceptualize universals of morality, even when those universals may be in conflict with local standards that would have been adhered to in the conventional level, the larger goal being principled rather than reactionary or instrumental morality (Dawson).

Measurement of Moral Reasoning

The study of moral reasoning development has been a long one, and so several researchers have developed tools with which to parse the developmental progress of individuals. Kohlberg first modified Piaget’s cognitive development interview method and developed the Moral Judgment Interview (MJI). In this measure an interviewee responds in a guided interview to a variety of moral dilemmas. In practice, however, the MJI is so laborious and thorough in its conduct and scoring that it has been relatively abandoned by researchers (Gibbs, Basinger & Grime, 2003). To promote and ease further research with larger sample sizes than can be had with interview modality, Rest (1990) developed the Defining Issues Test (DIT) and its successor the Defining Issues Test 2 (DIT2).

Similarly to Kohlberg’s MJI methodology, the DIT and DIT2 present a respondent with moral dilemmas, after which they rank order a variety of statements in relevance to, and agreement with the moral implications of the dilemmas. Even more recently, the Moral Judgment Test (MJT) was developed by Lind (2014). Lind developed this measure to assess cognitive as well as emotional aspects of moral reasoning through exposure to moral dilemmas followed by responses affirming or disaffirming statements associated with different degrees moral reasoning development.

The measures developed thus far have been very successful at more and more finely parsing the underlying influences on moral development, prime amongst them: education, gender, work-experience, and age. Education has been found to be a strong—though much debated—factor in moral reasoning development. Rest and Thoma conducted a longitudinal study, demonstrating a positive linear relationship between education and moral development (Bouhmama, 2001). The moral reasoning of those who attended college has been found to continue to increase while the moral reasoning development of those who did not attend college remained virtually the same (King and Mayhew, 2002). Gender difference findings typically have suggested that females attain higher levels of moral reasoning development than comparable males (Rest et al., 1999).
Work experience has been shown to be positively related to moral reasoning development, as well as the mitigation of gender effects as work experience increases (Franke, Crown, & Spake 1997; Windsor & Cappel, 1999). Age presents a greater source of controversy in that some results suggest that moral reasoning development increases with age (Gibbs & Widaman, 1982; Nisen & Kohlberg, 1982; Rest, 1983; Trevino, 1992) while some suggest a decline in moral reasoning is the more common result of aging (Rest, Davison, & Robbins, 1978). While the refinement of these measures and their expanded use have provided much insight into broad human development of moral reasoning, a practical weakness of the extant measures is their lack of moral dilemmas congruent with those faced by the practitioners of business.

It is easily conceivable and highly defensible to postulate that humans parse their moral reasoning by functional dimensions, given that rules of morality vary greatly between areas of human activity. For example, one would be severely reprimanded for lying as a child to their parent, but bluffing as a form of deceit is commonly practiced and revered in the game of poker. Representing someone else’s work as one’s own would be called plagiarism if a student were caught by a professor in doing so, however the practice of “appropriation” (the significant implementation of unaltered images and works of others for new works of art) is common practice in the arts, famously in Andy Warhol’s Campbell’s Soup pop-art silkscreens and Duchamp’s presentation of bottle racks (not manufactured by himself) as “ready-mades” in the French Salons. Given that humans separate their moral reasoning by functional area, it is entirely plausible that measures of moral reasoning that only focus on moral reasoning outside the arena of business may never accurately assess an individual’s moral reasoning within that arena.

HYPOTHESES

In order to test both the internal reliability and the external validity of the B-MRT as a measure of moral reasoning development, the authors of this study proctored it alongside the DIT2 and MJT for all participants in a Latin square rotation so that each measure preceded and was followed by each other measure for an equivalent number of participants. If the B-MRT is a sound measure, it should demonstrate significant inter-item reliability. Hence the null hypothesis has been developed:

\[ H_{01}: \text{Items in the B-MRT will not be reliably interrelated in such a way that the elimination of any individual one should improve upon the measure overall and the measure will have a low overall reliability as measured by Cronbach’s alpha.} \]

With 55 items being intercorrelated the chances of rejecting the null are not in the favor of the instrument.

If the B-MRT is veridically measuring the moral reasoning development of participants, it should bear significant relationship to one or both of the standard measures of moral reasoning, the DIT2 and MJT. This experimental hypothesis is stated positively as follows:

\[ H_2: \text{Scores of participants on the B-MRT will be significantly and positively correlated to their scores on one or both standard measures of moral reasoning, the DIT2 and MJT.} \]
It is anticipated that scores on the B-MRT may correlate more significantly with one measure than the other as the two measures themselves have not always been found to correlate well—or at all—with one another (Rest, Thoma, & Edwards, 1997).

Lastly, if the B-MRT is validly measuring moral reasoning development, we may expect that it should be able to distinguish between the influences of other predictors of moral development such as education, gender, work-experience, and age. Thus, the following experimental hypothesis has been developed:

H3: Variance of scores of participants on the B-MRT should be significantly predicted by one or more of the established factors in moral reasoning development—education, gender, work-experience, and/or age.

METHODOLOGY

The Business Moral Reasoning Test (B-MRT)

In order to better enable educators and researchers to assess—and eventually to improve upon—the moral reasoning development of business practitioners, the authors of this study have undertaken efforts to develop a tool that focuses specifically on dilemmas inherent to the arena of business. The authors, measure, the B-MRT is quite similar to both the DIT2 (Rest, et al., 1999) and the MJT (Lind, 2014). It contains a series of five moral dilemmas, each followed by eleven items that suggest a particular moral view of the primary actor in the dilemma and solicit the respondent to agree or disagree with each.

The moral dilemmas in the B-MRT developed from actual news events and business cases include the following:

1. An individual working for Housing and Urban Development (HUD) who supports the development of a project due to a strong conflict of interest, but with the implication of preserving quality of life for tenants of the project over the legal standards regarding disclosure;

2. An executive who decides to pursue the payment of bribes to local officials in China that will give their company equal footing competitively in a field of other companies who are presumably doing the same thing in violation of the Foreign Corrupt Practices Act (FCPA);

3. A Chief Legal Officer who decides to continue the extant, though informal, practice of bribing local militias to ensure the delivery of life saving drugs the company sells—as well as the safety of their drivers—in localities controlled by these militias;

4. A business owner deciding whether or not to continue the employment of an illegal immigrant whose family is dependent upon the income she earns, though the employment is in violation of federal law;

5. A medical practitioner whose prior knowledge of an employee includes that individual’s prescription drug addiction and must decide whether to maintain the individual’s employment given the risks of relapse and the violation of confidentiality to reveal such knowledge to their organization.
Much as with the DIT2 and MJT, each moral dilemma presented in the B-MRT is followed first by a question of either agreement or disagreement with (on an eight point scale from Strongly Agree [-3] to Strongly Disagree [+3]) the actions taken or expected to be taken by the primary actor in each dilemma. For example, in the case of the executive debating using a contractor who bribes for business in China, “Would you agree or disagree with using a contractor to obtain the licenses necessary to operate in China?” This item is followed by five arguments in favor of that action with which respondents are asked whether they accept or reject such arguments (on a nine point scale from Strongly Accept [-4] to Strongly Reject [+4]). From the same example case, some examples of these items are: 1) “Since Oswald is not directly supplying the cash to the agent, or paying the city or district for the license, there is no violation of the Foreign Corrupt Practices Act.”, 2) “Obtaining the additional license and increasing profits for shareholders is more important than worrying about what an agent might or might not have to do to obtain a license.”, and 3) “Having a license to do business in China will help the Chinese economy by providing more jobs for the local people.” These five items are followed by five arguments against the action, with participants again asked whether they accept or reject such arguments (again, on a nine point scale from Strongly Accept [-4] to Strongly Reject [+4]). From the same example case, some examples of these items are: 1) “If Oswald knows the agents might have to pay for license, it is still a violation of the Foreign Corrupt Practices Act despite the fact that the cash is not paid directly to the city or districts by Pegasus International.”, 2) “Pegasus should figure out how to expand their market share in countries that don’t require bribes in order to operate. This would allow them to increase their revenues without violating any laws.”, and 3) “If Pegasus pays these bribes they are contributing to behavior that perpetuates a society where those who hold positions of power can suppress those who do not.” Items were written with an eye toward the intersection of development found between applying local standards (e.g., legal regulations, precedents, common practices) and developing higher-order universal standards (e.g., protection of life, preservation of dignity and opportunity) in conflict with those local standards. This method very closely mimics the modus operandi of both the DIT2 and the MJT.

Participants and Data Collection

The authors of this study solicited voluntary participants for course credit at a small Midwestern university. Of those solicited, 159 completed all procedures and so were included in the study’s analyses. Of these 159, 75 were graduate students in the MBA and Masters of Accounting programs, 83 were undergraduates in a variety of business majors. The average age of participants was 27.89, and they possessed an average work experience of 10.76 years.

All participants were administered the B-MRT, DIT2, and MJT as well as a brief demographic questionnaire. Participants were instructed on and asked to complete the measures, and then answer demographic questions on a separate measure giving their age in number of years, indicating their level of education as undergraduate junior, undergraduate senior, or graduate student, identifying their academic major as accounting, business or other with space available to write in a response, and identifying their work experience and present field of work.
RESULTS OF THIS STUDY

All statistical tests were conducted using SPSS version 18, including the reliability measures conducted on the B-MRT. These tests of reliability showed an extraordinary internal reliability, Cronbach’s $\alpha = .916$, $n = 158$. All 55 items showed very high contributions to reliability such that none would benefit the scale’s overall reliability by their exclusion. By merit of its surprisingly high internal reliability, the B-MRT seems to have invalidated $H_0$ that it would not be reliable.

A reliable scale is only productive if it is measuring what it purports to measure and so correlational tests were conducted between participants’ B-MRT scores (calculated as the aggregated average of directionally aligned item scores), DIT2 scores (calculated by the Office for the Study of Ethical Development, the current owners of the DIT2 measure), and MJT scores (calculated using Lind’s [2014] c-score method outlined in his white papers and on his website). The results suggest that the B-MRT is indeed validly measuring moral reasoning development as it correlated significantly and positively with the MJT scores, $r = .219$, $p < .01$. Interestingly, but not entirely surprisingly, the B-MRT scores did not correlate with the DIT2 scores, $r = .057$, ns, which themselves correlated even less with MJT scores, $r = .011$, ns.

To further explore this relationship, we probed these correlations by controlling for the influential factors of work-experience and age. While correlations between the B-MRT and MJT scores remained significant and barely diminished, $r = .205$, $p < .01$, correlations between the B-MRT and DIT2 as well as the DIT2 and MJT remained nonsignificant, $r = 0.049$, ns, and $r = 0.015$, ns, respectively. These results taken together would suggest positive support for $H_2$, that the B-MRT is significantly correlated to MJT scores and therefore validly measuring the same construct of moral reasoning embodied in that popular measure.

In a last wholesale analysis to probe the external validity of the B-MRT, a MANOVA (Multivariate Analysis of Variance) was conducted, measuring the effect of education (graduate or undergraduate), gender (male or female), age (continuous, number of years), and work-experience (continuous, number of years), as well as the interaction of gender and work experience as some investigations have shown an interaction of these variables in particular, simultaneously on B-MRT, DIT2, and MJT scores. The general linear model for this analysis being:

$$B\text{-MRT (Avg.), DIT2 (n2), MJT (c)} = b_0 + b_1\text{Work Experience} + b_2\text{Age} + b_3\text{Gender (Male/Female)} + b_4\text{Education (Graduate/Undergraduate)} + b_5\text{Gender*Education}$$

The advantage of such a procedure is that it reduces statistical power compared with individual tests predicting each measure separately by including intercorrelations between the dependent variables themselves as a factor. In this analysis, we were surprised to find that the B-MRT as the only measure in the study capable of distinguishing the influence of education on variance in moral reasoning. While all other relationships remained nonsignificant as predictors of score on the measures of moral reasoning, education strongly accounted for variance in B-MRT scores, $F(1,151) = 6.316$, $p = .013$, but not the DIT2, $F(1,151) = 0.599$, ns, or the MJT, $F(1,151) = 1.429$, ns.
DISCUSSION

The present results suggest that the B-MRT is indeed a reliable and externally valid measure of moral reasoning development, novel in utilizing business specific dilemmas compared with the popular extant measures of moral reasoning, the DIT2 and MJT. Its internal reliability surprised even the authors, suggesting that any findings relevant to external validity apply to all dilemmas and items generated for it.

The significant correlation of B-MRT scores to those on the MJT suggests that the B-MRT is measuring moral reasoning development validly. The nonsignificant correlation to DIT2 scores as well as the nonsignificant correlation of MJT scores to DIT2 scores suggests that the B-MRT is tapping into elements of moral reasoning not captured in the DIT2, elements that also do not overlap between the MJT and DIT2. This is not unexpected as other studies have found that the DIT2 and MJT measure different aspects of moral reasoning development (Rest, Thoma, & Edwards, 1997). It is entirely plausible that the B-MRT’s isolated focus on business specific dilemmas differentiates it significantly from the DIT2. Further testing will be needed to refine our understanding of this relationship.

IMPLICATIONS, LIMITATIONS, AND RECOMMENDATIONS

The fact that the B-MRT was the only measure that’s scores were significantly predicted by any of the typical predictors of moral reasoning development—education, gender, work-experience, and age—strongly suggests that it is a sensitive measure of moral reasoning development, able to capture variance not embodied in DIT2 and MJT scores. The significant prediction of B-MRT scores tells a further heartening story that perhaps, given the business nature of all participants’ education, business education is cultivating the moral reasoning of students within the business realm of decision making. Because the sample is restricted to students at one institution and is parsed only by current undergraduate or graduate student identity, further testing will be needed to flesh this relationship out and see if it remains reliable across institutions, disciplines, and levels of education. Furthermore, research will be needed to assess the relationship between B-MRT scores and ethical decision making in other business specific contexts. The authors of this study are at present conducting an investigation comparing B-MRT scores with scored responses to an ethics case from the American Institute of CPAs (AICPA) ethics training materials in order to further validate the relationship between moral reasoning development in business specific contexts and ethical decision making within the same.

CONCLUSION

The authors of this study cannot deny that the state ethics in business presents a quandary for the populace and educators in particular. What we as the community of educators and researchers in business can do is attempt to better measure the ethical comprehension and habits of humans, specifically as it pertains to business decisions, and, through this measurement, attempt its improvement. The development and validation of the B-MRT is a strong first step in this direction, but true societal change will be a product not only of quality measurement, but also bravery and steadfastness in its use and the implementation of changes suggested by its outcome.
REFERENCES

http://www.aacsb.edu/~media/AACSB/Docs/Accreditation/Standards/2013-business-standards.ashx


*About the Authors:*

**Nathan Steele** is Associate Professor and Chair of Management at University of Illinois Springfield. Dr. Steele’s research interests are in the area of the impacts of social psychology on business practices.

**Leonard Branson** is Professor and Chair of Accountancy at the University of Illinois Springfield. Dr. Branson’s research interests are in areas related to behavioral accounting and curriculum development.