Journal of Educational Psychology
2009, Vol. 101, No. 1, 233–247

0022-0663/09/$12.00 DOI: 10.1037/a0013381

Understanding the Relationship Between Moral Judgment Development and Individual Characteristics: The Role of Educational Contexts

Yukiko Maeda
Purdue University

Steve J. Thoma
University of Alabama

Muriel J. Bebeau
University of Minnesota

This study examines variation in moral judgment level as measured by the Defining Issues Test 2 (DIT-2; J. R. Rest & D. Narvaez, 1998), based on individual demographic information and educational contexts. Individual DIT-2 scores and demographic information were obtained from the archived data sets housed at the Center for the Study of Ethical Development. The information on educational contexts was obtained by surveying the researchers who initially gathered the data. Hierarchical linear modeling was used to analyze DIT-2 data from 7,642 individuals and 65 institutions. Findings indicate that average DIT-2 scores vary significantly among institutions. Further, the authors found that the strength of the relation of DIT-2 scores to gender and English language status is significant and consistent across contexts, whereas the strength of the relation to educational level and political identity varies across contexts. These findings indicate that information about the individuals' educational context as well as the broader regional contexts should be taken into account to understand variation in individual levels of moral judgment. The findings can also be used as an interpretive guide for DIT-2 users.

Keywords: moral judgment, Defining Issues Test, higher education, educational contexts

Since its inception in the early 1970s, the Defining Issues Test (DIT) of moral judgment development has been used to study young adults enrolled in institutions of higher education. Driven in part by an interest in the conditions under which moral judgment develops, the study of college and graduate students soon expanded to include individual characteristics, institutional types and mission, and even region of the country (King & Mayhew, 2002; Pascarella & Terenzini, 1991, 2005; Rest, 1986). With the development of the DIT-2 (Rest & Narvaez, 1998), further attention was given to student characteristics through the inclusion of items soliciting specific information about each participant. These data have been archived by the Center for the Study of Ethical Development (CSED), and there now exists a large and diverse sample of students that can be used to study the joint effects of individual characteristics and institutional contexts, as well as larger community contexts, in young adult populations. Using statistical methods that estimate both individual and within context effects, this study attempts to explain variation in college student moral judgment development across region and university type as well as individual characteristics.

The extensive literature on college students—and to a lesser extent, graduate students—is primarily focused on why moral judgment development accelerates in these age groups. Specifically, it has been often noted in the DIT literature that late adolescence through early youth is associated with the transition from a conventional to a postconventional view of morality (Rest, Narvaez, Bebeau, & Thoma, 1999b; Thoma, 2006). A number of explanations for these findings have been suggested. For instance, researchers have wondered whether a shift from conventional to postconventional reasoning is due to an openness to change often associated with the college years. In this view, college is a time when youths are encouraged to explore their values and expectations for life, all of which illuminate moral and ideological issues (e.g., Rest & Thoma, 1985). Others suspect that students are influenced by the curriculum, both formal (e.g., coursework highlighting moral issues) and informal (the social environment that fosters a discussion of moral issues; e.g., Derryberry & Thoma, 2000). Still others have questioned whether the rapid shift in moral judgment development is best attributed to general maturation. Finally, some have wondered whether educational environments select for a particular type of individual who is interested in ideas and exploring values. Thus, in this view the relationship between moral judgment development and educational environments does not reflect any causal link between the two (see Rest, 1979a, 1986, for an extended discussion of these issues).

On the basis of more than 30 years of research, it does appear that educational environments are associated with growth in moral judgment development, and it is not simply an effect easily explained by maturation or selection. Longitudinal studies have consistently found upward movement on the DIT (e.g., King & Mayhew, 2002; Pascarella & Terenzini, 2005; Rest & Thoma, 1985). Cross-sectional trends in which age is controlled either statistically or through the

Yukiko Maeda, Department of Educational Studies, Purdue University; Steve J. Thoma, Department of Educational Studies in Psychology, Research Methodology, and Counseling, University of Alabama; Muriel J. Bebeau, Department of Primary Dental Care, University of Minnesota.

Correspondence concerning this article should be addressed to Yukiko Maeda, Department of Educational Studies, College of Education, Purdue University, 5119 Beering Hall of Liberal Arts & Education, 100 North University Street, West Lafayette, IN 47907-2098. E-mail: ymaeda@purdue.edu

design elements further support these longitudinal findings. Further, studies that focus on the integration of the individual into the college milieu have shown that an academic orientation is associated with moral judgment development (Deemer, 1989; Thoma & Ladewig, 1993). Indeed, all of the major reviews have reached the conclusion that—as King and Mayhew (2002) noted—the evidence to date "provides strong evidence that student participation in higher education is associated with gains in moral development during the college years" (p. 250).

Given the strong evidence that moral judgment development is associated with educational environments, research interests have shifted to the exploration of which particular environments promote growth (e.g., secular vs. religious schools; public vs. private; liberal arts vs. comprehensive) and for whom (e.g., White vs. ethnic minorities; men vs. women; King & Mayhew, 2002; McNeel, 1991, 1994). Unlike the data on educational environments, there is less of a consensus on whether these additional factors have a stable influence on moral judgment development.

Variation in Educational Environments

Educational environments vary in the degree to which moral issues are featured in the student's experiences, either within an institution, as measured by the students' program of study, or across institutions, by contrasting different types of higher educational environments based on their institutional mission (King & Mayhew, 2002). In spite of increasing attention to the linkages between variation in institutional types and the rate of growth in moral judgment development, research investigating the between-college effects on moral judgment development is rather limited (Pascarella & Terenzini, 2005).

Perhaps the first attempt to investigate such a link was made by McNeel (1994) in his meta-analysis of the variation in moral reasoning development among students in different types of colleges. McNeel found that students in private liberal arts colleges showed the largest positive gain in postconventional moral reasoning in their college years. Interestingly, he also found that students in Bible colleges showed the least positive gain in their college years. His findings were also consistent with findings by Good and Cartwright (1998). In their study, Good and Cartwright compared undergraduate students' moral judgment development at three different types of institutions: state university, Christian liberal arts university, and Bible university. They found that there were significant freshman-to-senior gains in DIT-1 postconventional (P) scores among students in the state and Christian liberal arts universities, whereas no significant gain was observed among students in Bible universities. McNeel's earlier work (McNeel, 1991) also indicated that whereas deep commitment to conservative Christian ideology may inhibit growth of moral judgment, this does not apply to Christian institutions in which education truly focuses on the liberal arts.

McNeel's (1991, 1994) and Good and Cartwright's (1998) findings were also supported by King and Mayhew's (2002) review of DIT studies conducted at higher education institutions. King and Mayhew noted that there is considerable variation in findings across studies on how institutional type functions to promote students' moral judgment development. However, it does appear that liberal arts institutions—both secular and religiously affiliated—are associated with the largest and most consistent gains on DIT scores. Exposure to the liberal arts curriculum is more central than the religious component to students' growth in postconventional moral thinking. Indeed, Pascarella and Terenzini (2005) synthesized findings from studies conducted in the 1990s and concluded that "traditional liberal arts education, combined with the unique social psychological environments of small liberal arts colleges, may be particularly conducive to fostering growth in principled moral reasoning" (p. 368).

The link between moral judgment development and student participation in different disciplines has been studied, but rarely provides a consistent pattern of findings (King and Mayhew, 2002; Pascarella & Terenzini, 2005; Rest, 1986). Often these studies are guided by the notion that students who are exposed to the social sciences and humanities should develop at a faster rate than their peers in more technical fields. That is, following a traditional cognitive development view of the process of growth and development (e.g., Rest, 1986), the more often students are exposed to a range of ideas about social institutions, social theory, and the good life, the more likely they are to grow in their understanding of sociomoral phenomena. The failure to consistently support this prediction is due, in part, to the rather broad and heterogeneous categories used to compare the different disciplines (Rest, 1986).

Variation in Educational Effects Due to Individual Characteristics

In addition to type of institution and different courses of study, researchers have also focused on the characteristics of the student, such as gender, political orientation, and ethnic background. Of interest to these researchers is whether different types of students experience the educational environments in different ways and thus show different patterns of growth in moral judgments.

Gender has a long history in moral judgment research (Walker, 2006). Although there is growing consensus that men and women are more similar than different on measures of moral judgment development such as the DIT (e.g., Thoma, 1986), it may be that gender interacts with different institutional settings in promoting growth. For example, White (1999) found a significant difference favoring females in the average moral judgment level between males and females at every level of rank within the U.S. Coast Guard. On the other hand, Wilson (1995) found no statistically significant differences between male and female registered nurses. In addition, Bebeau (2002) noted that gender differences in P scores that favor females have been observed in medicine, veterinary medicine, and law (e.g., Landsman & McNeel, 2004; Self & Baldwin, 1998; Self, Pierce, & Shadduck, 1995). Bebeau (2002) also indicated that the magnitude of gender difference reported in studies of professional students is consistently larger than that in nonprofessional graduate students.

In addition to gender, political identity is an individual characteristic that has been studied in the context of educational environments. Viewed in terms of liberalism versus conservatism, political identity has consistently been linked to moral judgment development (Rest, Narvaez, Bebeau, & Thoma, 1999b). Although the explanation for why political liberals achieve higher scores than conservatives is somewhat controversial (Emler, Resnick, & Malone 1983; Rest, Narvaez, Bebeau, & Thoma, 1999a, 1999b; Thoma, Barnett, Rest, & Narvaez, 1999), DIT researchers consider political identity and moral judgment to be independent constructs

(e,g., Crowson, DeBacker, & Thoma, 2007; Rest & Thoma, 1985). They view the relationship between political and moral judgments as the result of different conceptualizations of social cooperation. For example, individuals who prefer the maintaining-norms moral schema value the existing social norms as the basis of a cooperative society and are suspicious of any attempt to change the status quo. On the other hand, people who prefer the postconventional moral schema are viewed as more willing to question existing social norms and more open to social change. In the current political configuration, a view that prioritizes social norms as the basis of social cooperation is consistent with a politically conservative perspective. By contrast, a liberal perspective is more closely aligned with a willingness to challenge and modify social conventions (see Rest, Narvaez, Bebeau, & Thoma, 1999b, for an extended discussion of the theoretical relationship between moral judgment and political orientation). Most of the data supporting the link between political orientation and moral judgment have been collected with college student populations. However, little is known about the interaction between political orientation and growth during the college years. It may be that liberals have an overall worldview that is more receptive to the range of ideas presented within the formal curriculum. For instance, in related work, children who are raised in households with liberal parents were found to develop a political language that seemed, in part, to foster the growth of moral judgments (Marx, 2005). Further, it is not known whether the rate of change in moral judgment differs for self-proclaimed liberals or conservatives. See Rest et al. (1999b) for a complete discussion of these issues.

Variation in Educational Effects Due to Larger Context Effects

In addition to type of institution and individual characteristics, the larger context in which the institution is located may be related to the development of moral judgment. Early in the development of the DIT, Rest (1979a) reported that there were regional differences in average DIT P scores across the United States. In this preliminary work, samples from two colleges in Georgia and Virginia produced significantly lower average P scores than the overall college average. Since this earlier work, there have been a number of studies that have replicated the basic finding of regional difference (e.g., Rest, 1986). DIT researchers have tentatively reasoned that the lower average P score in southern parts of the United States may be due to the conservative intellectual environment. In this view, the southern states are associated with a fairly uniform conservative political climate that does not lend itself to debates over social–moral issues. Thus, students from this region may not be as likely to experience contrasting views within their peer groups in college and in the larger community.

Limitation of the Current Research Base

It is important to note that the findings described above are mainly derived from studies using the original DIT (DIT-1; Rest, 1979b). These data were collected across cohorts from the 1970s to the early 1990s, which may or may not accurately reflect current college student populations. The Defining Issues Test 2 (DIT-2; Rest & Narvaez, 1998) was first introduced in 1998 as an updated version of the original DIT. The DIT-2 presents alternative stories and items to the original DIT. The changes relate to more modern contexts for stories and language considered to be dated (Rest, Narvaez, Thoma, & Bebeau, 1999). In addition, demographic questions (i.e., gender, highest educational attainment, political orientation, and language status) were added to the response form for collecting information directly from respondents. Because the DIT-2 has better psychometric properties, reliability checks, and summary indices (Rest et al., 1999), the DIT-2 has now superseded the DIT-1, and efforts have been initiated to replicate earlier findings based on the DIT-1.

The earliest work designed to assess the DIT-2 and relationships to individual demographic characteristics was conducted by Rest et al. (1999) using a total of 200 subjects from different educational levels. Similar to the findings with the original DIT, the DIT-2 scores increased as age and educational level increased. Recently, Bebeau, Maeda, and Tichy-Reese (2003) attempted to establish more extensive DIT-2 norms based on the archived data compiled by the CSED. The data consist of 10,870 subjects who completed the DIT-2 and demographic questions. This study verified that the relationship between DIT-2 scores and demographic characteristics, such as educational level, gender, and political orientation, was consistent with most of the findings for the original DIT. Specifically, Bebeau et al. (2003) replicated the finding that moral judgment scores increase with level of education and are slightly higher for females (although this gender difference decreases with level of education). Additionally, the relationship between political orientation and the moral judgment level was evident. Liberals scored higher on postconventional items than their conservative peers. Further, this difference in DIT scores by political identity became more pronounced following the sophomore year of college (Bebeau & Thoma, 2003) and continued to be evident through graduate education.

Although individual and educational contexts seem to be jointly associated with moral judgment development, no study has been conducted to examine statistically the relation of individual and institutional characteristics to individual growth of moral judgment measured by the DIT. With this in mind, and focusing on individuals in higher educational institutions in the United States, the purpose of this study is to extend the work done by Bebeau and Thoma (2003) by taking educational contexts into account. More specifically, the current study is an attempt to jointly assess relationships between institutional characteristics, regional differences, and growth in moral judgment as measured by the DIT-2. To address this question, we first asked whether average moral judgment levels vary across separate data sets obtained from a variety of educational contexts. That is, as a precondition to an analysis linking individual and regional characteristics with DIT-2 scores, the data should indicate significant variation across studies that include different characteristics. Three additional questions were then addressed: whether DIT-2 scores vary by individual characteristics (e.g., gender and political identity) or educational level; whether educational contexts relate to individual DIT-2 scores; and whether the relationship between individual DIT-2 scores and individual characteristics varies by educational context. Thus, the study can be viewed as a meta-analysis in which we assess whether individual characteristics and settings account for significant variation in DIT-2 scores, either individually or in combination.

Assessing the Link Between DIT Scores and Educational Contexts

Most previous DIT studies have used statistical techniques, such as the independent samples *t* test and analysis of variance (ANOVA), to investigate the differences between the average DIT scores of different groups, potentially ignoring other important characteristics that may jointly affect moral judgment level. This shortcoming has not gone unnoticed. King and Mayhew (2002) argued that the use of advanced statistical techniques was necessary to investigate the unique contribution of specific individual and institutional demographic characteristics on moral judgment. To that end, the current study used a two-level hierarchical linear model (HLM; Raudenbush & Bryk, 2002) in order to analyze DIT-2 data. HLM is often used in educational research because it is designed to account for the inherent dependency that exists in groups of individuals who share a particular context (e.g., students within a particular institution). That is, when individual data are nested within a context, dependency exists in these data because people in the same context share experiences unique to that context (e.g., Raudenbush & Bryk, 2002; Snijders & Bosker, 1999). Using traditional regression approaches for nested data is problematic because the dependency is ignored and can result in drawing invalid conclusions based on inaccurate results. HLM, on the other hand, is designed to take the dependency into account in analyses.

In addition, the application of HLM enables one to investigate questions that may be unanswerable with traditional regression models. For example, a researcher can investigate the relationship between group-level characteristics, such as religious affiliation (e.g., secular vs. religious schools), and individual-level outcomes, such as individual moral judgment. Finally, the ability to assess cross-sectional interactions is particularly important in the current study because it allows us to address questions about whether institutional type moderates the relationship between DIT-2 scores and individual characteristics.

Method

Data

Individual DIT data used for the current study were from archival data obtained from 1998 to 2005 at the CSED. The data coded for each individual included DIT-2 raw responses, as well as the standard summary scores: P score, Maintaining Norms (MN) score, Personal Interest score, and N2 score. These DIT scores reflect the degree to which the individual uses different types of schemas to make a moral judgment. The P score reflects an individual's preference for postconventional moral thinking, whereas the MN score reflects the degree to which an individual emphasizes the maintenance of societal norms. The Personal Interest score indicates the extent to which an individual's own personal interests or relationships guide his or her moral thinking. The N2 score is a new index that tends to show stronger construct-related validity than the P score. As described in Thoma (2006), "the N2 score uses the P-score as its starting point and then adjusts the P-score based on participants' ability to discriminate between P [Postconventional] items and lower stage items" (p. 80). Also included was demographic information for each individual, such as age, gender, highest educational level, political identity, citizenship status, and English language status. The demographic information is also available in DIT-2 archived data.

Information on the institutional setting was obtained by surveying researchers who requested the DIT-2 scoring service of the CSED between 1998 and 2005. By inspecting these files we were able to identify a total of 231 researchers who had submitted DIT-2 responses to the scoring service. From this pool of 231 researchers, we were able to locate the current addresses of 183 DIT-2 users. Of these 183 researchers, 97 provided information about their test sites, yielding a response rate of 53.0%.

To estimate the possibility of a selection bias in our working sample, we assessed whether the characteristics of the responders were different from those who did not respond. We did this in two ways. First, we compared the two groups in terms of geographic location and type of institution. Inspection of these lists indicates that there was no systematic difference in institution type and location within the United States. Second, we noted that nonrespondents were more likely to be one-time users of the DIT and no longer in the field. One possible interpretation of this observation is that the nonrespondents overrepresented graduate students who used the DIT-2 for their thesis or dissertation.

The working data set of 97 was further reduced to 65 unique samples collected from higher education institutions in 28 states in the United States. Data sets were eliminated for these reasons: When DIT-2 data sets were gathered from multiple sites by a single researcher, all but the largest sample were excluded unless the additional DIT-2 data set could be unambiguously attributed to a distinct setting. DIT-2 data sets obtained from outside of the United States were excluded, as they would confound potential cultural differences in the meaning of the variables as well as the links between them (see, e.g., Rest et al., 1999). More practically, the number of data sets collected outside the United States was not sufficient to investigate national differences. Finally, most researchers in non–English-speaking countries used some type of translated version of the DIT-2, the majority of which have limited psychometric information and empirical support. In sum, 65 of the identified separate DIT-2 data sets from CSED were merged. Then the data were rescored to identify any bogus DIT data, using the reliability check index designed to identify individuals who provided either random or careless responses. The responses of these identified individuals were also excluded from the current study. The rejection rate based on the reliability checks was 5.9% of the data; thus, the final data set for the current study consisted of 7,462 individuals from 65 institutions. The sample size within an institution ranged from 7 to 863, and the median sample size was 68.

Measures

Defining Issues Test, Version 2 (Rest & Narvaez, 1998). The DIT-2 consists of five stories that present moral dilemmas (see Appendix A for a sample story from the DIT-2.) After reading each story, a respondent is asked to complete three different tasks. The first task is to select the subsequent action that a respondent thinks the character in the story would take from three listed options; for example, for the story in Appendix A, there are the following choices: (a) *should take the food*, (b) *can't decide*, (c) *should not take the food*. The DIT-2 contains 12 statements of possible reasoning for making a decision about selecting the succeeding action (e.g., *Is Mustaq Singh courageous enough to risk getting caught for stealing?*). The second task is to rate the

importance of each of these 12 statements of reasoning on a 5-point scale (i.e., from *no importance* to *great importance*), based on his or her own moral schema. The final task is to select 4 of the 12 statements deemed most important in making a decision and rank them as *most important*, *second most important*, and so on.

The DIT-2 schema scores are computed on the basis of these four ranked statements. The calculated schema scores are expressed as the ratio of an individual's obtained score for each schema to the total possible score; thus, the scores range from 0 to 99. Of the various DIT indices, the postconventional (P) score, indicating the relative importance of postconventional reasoning, is the most widely used index in DIT research. Thus, the P score serves as a primary outcome variable, whereas the study also reports the results with the N2 scores, which is a newly developed index that tends to show greater construct validity evidence than the P score (Rest, Thoma, Narvaez, & Bebeau, 1997).

The reliability of the P score with the current data is 0.639, which is lower than the reliability of 0.74 reported by Rest et al. (1999). The lower reliability observed here can probably be attributed to the fact that the sample used in the current study is more homogeneous than that used by Rest et al. with regard to educational level. The reliability of the N2 score with the present data is 0.732.

The DIT-2 also contains six questions that inquire about individual characteristics, such as the highest attained education level (12 categories range from *Grade 1 to 6* to *doctoral degree*), gender (*male* or *female*), political orientation (5 rating categories range from *very liberal* to *very conservative*), English language status (2 categories indicating whether English is primary language or not), citizenship (2 categories indicating whether the individual is a U.S. citizen or not) and age (a continuous variable ranges from 0 to 99; the median age of the current sample is 22.0). The current study does not investigate the relationship between U.S. citizenship status and the DIT-2 score, as less than 4% of the respondents indicated that they were not U.S. citizens. In addition, the age effect was not investigated with the current study, because the variables indicating age and education level are highly correlated ($r = .742$).

Questionnaire for educational contexts. A questionnaire was developed for use in the current study to gather information with regard to DIT-2 respondents and the educational contexts where individual data were obtained. The questionnaire contains items asking about the location, type of the institution (i.e., public or private), and religious affiliation of the higher education institution, and academic major of respondents.

Variables

Individual characteristics. As previously mentioned, information regarding individual characteristics (i.e., education level, political orientation, gender, and English language status) was obtained from each respondent through the standard set of demographic questions collected during DIT-2 administration and archived with his or her DIT-2 schema scores. Information on these individual characteristics in the DIT-2 archived data was first recategorized for use in the statistical analyses. More specifically, each individual's highest education level was represented as an ordinal scale with five levels. The lowest education level was *college freshman*, and the highest level was *graduate student*. Political orientation was also represented as an ordinal scale, using each individual's identification of his or her own political views: −1 for *liberal*, 0 for *neither liberal nor conservative*, and 1 for *conservative*. Gender and English language status were dummy coded with a code of 1 for male and 0 for female, and with a code of 1 for native English speaker and 0 for nonnative English speaker.

Educational contexts. Institutional characteristics, such as religious affiliation (i.e., religious vs. secular schools) and academic major (i.e., medical school vs. business major vs. primarily arts and sciences), were selected as being indicative of educational context. Based on the information obtained through the questionnaire for educational contexts, religious affiliation of institutions was identified for each of 65 data sets. Religious affiliation was categorized dichotomously, with a code of 1 for a religious school and 0 for a secular school.

Next, the data sets were categorized for academic major based on information obtained by surveying the researchers who collected the data. Most of the 65 data sets were gathered in classroom settings representing a variety of academic majors. In fact, 5 of the 65 data sets were collected at medical schools and consist of 999 medical students. Similarly, 14 of the 65 data sets were gathered at a business school or from students in a business major and consist of 1,206 students. Compared with groups of individuals in other data sets, representing a variety of academic majors, these groups were unique and homogeneous in terms of their academic experiences. Therefore, to capture the difference in their academic experiences, two dummy variables were created. For the first dummy variable, the code of 1 was assigned if the data set was collected at a medical school and 0 if otherwise. For the second dummy variable, the code of 1 was assigned if the data set was collected at a business school or from students majoring in business and 0 if otherwise.

The study also employed two variables that measure the broader environment (i.e., political climate—politically conservative vs. liberal—and the region where the data were collected.) The 65 data sets were initially classified into four categories on the basis of the U.S. Census Bureau's census regions (http://www.census.gov/geo/www/us_regdiv.pdf). Subsequently, the data sets were reclassified into two groups (southern region vs. nonsouthern regions) for the succeeding HLM analyses.

Sixty-five data sets were also dichotomously categorized for the political climate in the state where an institution was located. Political climate was identified based on the results of the 2004 presidential election. States where the Republican candidate won a plurality were categorized as politically conservative (coded 1); states where the Democrat won a plurality were labeled politically liberal (coded 0).

Because there is almost no variation in the political climate in the southern region (i.e., only one data set obtained from the southern region is identified as politically liberal), the combined variable for political climate and region was created for the analysis with the HLM. Specifically, three coding categories were developed: southern conservative states, nonsouthern conservative states, and liberal states.

Finally, an ordinal variable (ranging from 1 to 5) indicating academic ranking of the institutions was created, based on the Carnegie Classification of Institutions of Higher Education (http://www.carnegiefoundation.org/classifications) developed by the Carnegie Foundation for the Advancement of Teaching (2007). Sixty-five institutions were categorized in one of the five ordered categories using the published Carnegie rankings. The lowest

category of the academic rank contains the institutions that offer associate degrees, such as community colleges; the second lowest category contains the institutions that offer mainly baccalaureate degrees; the third category contains master's-degree-granting institutions; the fourth category contains doctoral-degree-granting non–research-oriented institutions; the fifth category, which is the highest academic rank, contains the doctoral-degree-granting research institutions, including medical institutions.

Analyses

First, a series of descriptive analyses was conducted to investigate the differences in the average P and N2 scores as functions of individual and educational contexts. Independent samples *t* tests or one-way fixed effects ANOVAs were conducted to examine whether a significant group difference existed among the defined groups. Second, the data were analyzed using two-level HLM, modeling individuals at the first level and educational contexts at the second level. Three different HLMs were fitted to investigate the research questions. First, the one-way ANOVA with random effects (see Appendix B) was fitted to investigate whether there was a significant variation in the average DIT-2 scores across 65 data sets obtained from different educational contexts.

Second, the random coefficient model (see Appendix B) was fitted to the data to investigate whether the moral judgment level varied across individuals according to demographic variables, such as educational level, gender, political orientation, and English language status. The model contains individual demographic characteristics in the within-group model to examine the effects of these characteristics on the DIT-2 scores. In other words, the model includes information on the highest level of education, language status, political orientation, and gender to explain the variation in the DIT-2 scores among individuals. The random coefficient model also has five between-groups models. The intercept and slopes of the within-group model served as the dependent variables of between-groups models. No educational context was modeled in the between-groups models.

The HLM fits a unique regression line to predict the DIT-2 score of individuals within a group with these common predictors for each of the groups simultaneously. Therefore, this random coefficient model not only estimates the size of the relationship for each individual demographic variable, but also examines whether the magnitude of these relationships on the DIT-2 scores is the same among different educational contexts.

Finally, the random intercept and slope model was fitted to the DIT-2 data. This model contains the variables for educational contexts in the between-groups model to explain contextual variation in the strength of the effect of individual demographic characteristics, as well as the variation in the average P scores. With the variables available for the current study, the model that was best able to explain the variation in DIT-2 scores is reported in the Results section. An alpha level of .05 was selected to examine the statistical significance for HLM analyses.

Results

Preliminary Analyses

Because the P score and N2 score are highly correlated ($r = .91$), the results obtained with these scores were similar for most of the analyses. However, due to current DIT usage patterns, we report our findings using both the currently recommended N2 summary score (e.g., Rest, 1998) and the more traditional and widely known P score.

Table 1 lists the descriptive statistics of the P score and the N2 score as a function of gender, language status, educational level, and political orientation. Although all individual characteristics are significantly related with DIT scores, the magnitude of the relationship is small and varies from .04 to .27. The strongest correlation is found between the individual education level variable and the DIT-2 scores ($r = .27$). As expected, this correlation and the descriptive statistics indicate that the P and N2 scores tend to increase as the education level increases.

The correlation between gender and the P score is $-.13$. This indicates that females (coded as 0) tend to get higher DIT-2 scores than males (coded as 1). The descriptive results of gender reveal that, on average, there is a 4-point difference between the P scores of females and males. The correlation between political orientation and the P scores is $-.14$, which means that the P score tends to be lower when individuals perceive themselves as politically conservative.

Table 2 reports the descriptive statistics between average DIT-2 scores and educational contexts. The results of one-way fixed effects ANOVA indicate that there is a significant difference in the average DIT-2 scores among groups for region: P score, $F(3, 61) = 3.97$, $p = .01$, N2 score, $F(3, 61) = 3.10$, $p = .03$; and political climate: P score, $F(1, 63) = 7.85$, $p < .01$, N2 score, $F(1, 63) = 6.41$, $p = .01$. The eta-square, indicating the effect size, for region is 0.16 for the P score and 0.13 for the N2 score. Similarly, the eta-square for political climate is 0.11 for the P score and 0.09 for the N2 score.

In addition, a statistically significant result was obtained using the political climate and region combination variable on group DIT-2 scores: P score, $F(2, 62) = 4.83$, $p = .01$; N2 score, $F(2, 62) = 3.69$, $p = .03$. The eta-square is 0.14 for the P score and 0.11 for the N2 score. By contrast, the results also showed that there are no statistically significant differences in the average P scores for the status of religious affiliation: P score, $F(1, 63) = 1.73$, $p = .19$, N2 score, $F(1, 63) = 1.75$, $p = .19$; academic major, P score, $F(2, 62) = 2.04$, $p = .14$, N2 score, $F(2, 62) = 2.42$, $p = .09$. Finally, because the academic rank variable is an ordinal variable, Spearman's rho was computed to indicate the magnitude of the relationship between institutional averages of P scores and academic rank of institutions. The relationship of institutional averages of P scores to academic rank is weak but statistically significant ($r = .25$, $p = .04$), whereas the relationship to N2 score is not statistically significant ($r = .21$, $p = .09$).

HLM Analyses

Estimating the variation in the average DIT-2 scores across institutions. The results of the one-way ANOVA with random effects for the P score and N2 score are reported in Tables 3 and 4, respectively. Results indicate that there is significant variation in the average P scores across institutions, $\chi^2(64) = 1{,}462.69$, $p < .01$. Thus, the first step in the analysis confirms that there is a significant amount of variance to explain by setting characteristics, such as the ones proposed in the current study.

The estimated overall average P score is 37.37 points, which is consistent with current DIT-2 norms (Bebeau & Thoma, 2003). From the result, we can also conclude that about 17.4% of the total

Table 1
Descriptive Statistics of the Postconventional (P) Score and the N2 Score as Functions of Individual Demographic Characteristics

	P score				N2 score			
Individual characteristics	*n*	*M*	*SD*	*r*	*n*	*M*	*SD*	*r*
Gender								
Female	4,401	39.02	15.88		4,399	38.45	15.76	
Male	3,128	34.98	15.55	−.13**	3,123	34.41	15.70	−.13**
Language status								
Nonnative English speaker	437	35.08	15.18		437	33.84	15.13	
Native English speaker	7,125	37.49	15.91	.04**	7,119	36.98	15.89	.05**
Education								
Freshman	1,616	30.98	13.97		1,611	30.09	14.51	
Sophomore	508	32.46	14.18		508	31.71	14.05	
Junior	850	35.59	15.65		850	34.53	15.87	
Senior	1,660	38.26	16.06		1,660	37.80	15.82	
Graduate	2,832	41.60	15.63	.27**	2,830	41.37	15.15	.28**
Political orientation								
Liberal	2,692	40.70	16.79		2,690	39.40	16.52	
Neither	1,783	35.94	15.56		1,780	34.92	15.73	
Conservative	2,984	35.42	14.75	−.14**	2,982	35.81	15.04	−.10**

** $p < .01$.

variation in P scores among individuals was due to the difference in test sites. Similarly, there is significant variation in the average N2 scores across test sites, $\chi^2(64) = 1{,}708.65$, $p < .01$. The estimated overall average N2 score is 36.80. About 20.2% of the total variation in N2 scores among individuals is attributable to institutional difference. In summary, the results indicate that there is significant variation in the moral judgment level among the institutions where DIT-2s were administered.

Do individual characteristics account for variation in the DIT-2? To address this first major research question, individual demographic characteristics (i.e., gender, education level, language status, and political orientation) were modeled to explain the within-group variation. The initial fitting of this model indicated that the random effects of gender and language status were not significant, indicating that there is no variation in the strength of the relationship between these two variables and DIT-2 scores among groups. That is, the effect of gender and language on DIT-2 scores does not vary across settings. Due to this finding, the model was refined by constraining the relationship between these two variables and the DIT-2 scores to be the same for all groups. The final estimation of fixed and random effects of the predictors with the random coefficient model is reported in Tables 3 and 4.

As expected, all modeled individual characteristics were found to be significant, even when all other effects in the model are held constant. For example, the language status variable was found to have the largest effect on P scores (4.75). This indicates that there is about a 5-point difference between the P scores of native English speakers (coded as 1 in the variable) and nonnative English speakers (coded as 0) whose educational level, gender, and political orientation are identical. The fixed-effect coefficient of gender is −4.40, meaning that on average, males (coded as 1) tend to score about 4.4 points lower on the P score than females (coded as 0) when their education level, language status, and political orientation are identical. Similar interpretations can be made of the other significant fixed-effect coefficients.

A deviance test for the model fit (Snijders & Bosker, 1999, p. 88) indicates that the random coefficient model for the P score fits better than the random-effects ANOVA model, $\chi^2(9) = 3{,}429.90$, $p < .01$. That is, compared with the result of the unconditional model, adding these four variables in the model leads to a reduction of the within-group variation in the P scores by 8.1%. Note that this finding also suggests that there remains significant unexplained variation in the P scores among individuals within an institution.

The result obtained from the N2 score is consistent with that obtained from the P score. The result of the deviance test for the N2 score indicates that the random coefficient model fits better than the one-way ANOVA with random effects, $\chi^2(9) = 3{,}437.46$, $p < .01$. The set of four individual demographic variables accounted for 8.4% of within-group variation in the N2 scores.

It is interesting that the result also indicated that there are significant variations in the effects of education level and political orientation on the DIT-2 scores and the intercept across settings. In other words, although there is a statistically significant relationship between educational level and DIT-2 scores, the magnitude of the relationship varies between institutions. Similarly, the magnitude of the significant relationship between political orientation and the P scores (or the N2 scores) varies across institutions.

Explaining the variations in the effects of educational level and political identity on DIT-2 scores. To explain the variation in the intercept and these slopes across institutions, institutional characteristics were included in the model. Initial fitting of the HLM model with the educational contexts showed that a dummy variable indicating the nonsouthern politically conservative region is not related to any of the variations in the above-mentioned effects among different educational contexts (i.e., the *p* values associated with the effect of nonsouthern politically conservative region on the intercept, education, and political identity slopes for the P and N2 scores ranged from 0.07 to 0.64 and from 0.10 to 0.53, respectively). Thus, the variable was excluded from the final model. Given that the set of dummy variables was created to explain differences in effects of education and political orientations across three geographic regions, these findings indicate that there is no difference in education and political orientation relationship with DIT-2 scores between nonsouthern politically conservative regions and other politically liberal regions. Similarly, a dummy variable indicating the business major

Table 2
Descriptive Statistics of the Postconventional (P) Score and the N2 Score as Functions of Educational Contexts

Educational context	*N*	P score			N2 score		
		M	*SD*	η^2	*M*	*SD*	η^2
Region*							
Northeast	10	35.04	7.25	.16	34.24	7.85	.13
Midwest	20	39.28	7.83		38.60	8.64	
West	14	41.70	8.44		40.97	8.83	
South	21	34.51	3.85		34.17	4.47	
Political climate*							
Liberal	29	40.30	8.28	.11	39.63	8.56	.09
Conservative	36	35.44	5.65		34.89	6.52	
Religious affiliation							
Secular	49	36.93	7.28	.03	36.28	7.80	.03
Religious	16	39.68	7.23		39.23	7.65	
Region and political climate, combined*							
Southern, conservative	20	34.09	3.43	.14	33.79	4.23	.11
Nonsouthern, conservative	16	37.13	7.36		36.27	8.53	
Other states	29	40.30	8.28		39.63	8.56	
Academic ranking of institutions							
Associate granting	2	29.11	1.98	[a]	27.14	4.43	[b]
Baccalaureate granting	9	36.14	5.34		36.04	4.88	
Master's degree granting	14	35.65	4.69		34.97	5.34	
Non–research-oriented doctoral degree granting	8	37.66	10.49		36.34	11.01	
Research-oriented doctoral degree granting	32	39.39	7.64		38.95	8.20	
Academic major							
Medical school	5	43.35	6.94	.06	43.44	6.90	.07
Business	14	35.80	6.26		34.69	7.18	
Other	46	37.53	7.46		37.01	7.83	

Note. Descriptive statistics were computed using group data set as a unit of analysis ($N = 65$).
[a] Spearman's rho is 0.25 with the *p* value of .04. [b] Spearman's rho is 0.21 with the *p* value of .09.
* $p < .05$.

was also removed from the final model because the variable was not related to any of the variations (i.e., the *p* values associated with the effect of business major on the intercept, education, and political identity slopes for the P and N2 scores ranged from 0.15 to 0.99 and from 0.12 to 0.85, respectively). This finding indicates that there is no moderating effect of the business major on the relation of the DIT score to education or to political orientation. This finding further indicates that there is no statistically significant difference in the education and political orientation relationship with DIT scores between students in the business major and other students who are not in either the business or the medical samples. Thus, the final model contains the dummy variable that distinguishes between medical students and all others, including business students.

Table 3
Summary of HLM Analyses of Unconditional and Random Coefficient Models for Variables Predicting Postconventional Score

	Unconditional model			Random coefficient model		
	Coeff.	*SE*	*t*	Coeff.	*SE*	*t*
Fixed effect						
Intercept	37.37	0.87	43.10	37.43	0.89	42.24*
Gender	—	—	—	−4.40	0.39	−11.23*
Education	—	—	—	2.11	0.29	7.18*
Language	—	—	—	4.75	1.01	4.68*
Political orientation	—	—	—	−1.93	0.28	−6.92*
	SD	*df*	χ^2	*SD*	*df*	χ^2
Random effect						
Within-group residual	14.59			13.99		
Intercept	6.69	64	1,462.69	6.86	57	1,397.67*
Education	—	—	—	1.06	57	101.27*
Political orientation	—	—	—	1.33	57	104.88*

Note. A dash indicates that the effect was not estimated. HLM = hierarchical linear modeling; coeff. = coefficient.
* $p < .05$.

Table 4
Summary of HLM Analyses of Unconditional and Random Coefficient Models For Variables Predicting the N2 Score

	Unconditional model			Random coefficient model		
	Coeff.	*SE*	*t*	Coeff.	*SE*	*t*
Fixed effect						
Intercept	36.80	0.93	39.54	36.81	0.95	38.93*
Gender	—	—	—	−4.83	0.41	−11.76*
Education	—	—	—	2.26	0.30	7.59*
Language	—	—	—	5.31	0.99	5.34*
Political orientation	—	—	—	−1.25	0.25	−5.06*
	SD	*df*	χ^2	*SD*	*df*	χ^2
Random effect						
Within-group residual	14.37			13.75		
Intercept	7.22	64	1708.65	7.36	57	1,651.47*
Education	—	—	—	1.12	57	96.74*
Political orientation	—	—	—	1.00	57	103.23*

Note. A dash indicates that the effect was not estimated. HLM = hierarchical linear modeling; coeff. = coefficient.
* $p < .05$.

The final model included variables indicating medical school (i.e., medicine vs. others), religious affiliation of schools (secular vs. religiously affiliated), academic rank with five ordered categories, and a dummy variable for the location of the institution (i.e., southern politically conservative states vs. others). All variables in the between-groups models are grand mean centered. Because there is no variation in the relation between gender and language status and the DIT-2 scores across institutions, no predictor was modeled in their between-groups model.

Within-group regression model:

$$Y_{ij} = \beta_{0j} + \beta_{1j}(\text{Gender}_{ij} - X_{\text{gender}\bullet j}) + \beta_{2j}(\text{Political Orientation}_{ij} - X_{\text{politics}\bullet j}) + \beta_{3j}(\text{Language Status}_{ij} - X_{\text{language}\bullet j}) + \beta_{4j}(\text{Education Level}_{ij} - X_{\text{education}\bullet j}) + r_{ij}$$

Between-groups regression model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(\text{Medicine}_j - X_{\text{med}\bullet\bullet}) + \gamma_{02}(\text{Religion}_j - X_{\text{reli}\bullet\bullet}) + \gamma_{03}(\text{Southern, conservative}_j - X_{\text{location}\bullet\bullet}) + \gamma_{04}(\text{Rank}_j - X_{\text{rank}\bullet\bullet}) + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}(\text{Medicine}_j - X_{\text{med}\bullet\bullet}) + \gamma_{22}(\text{Religion}_j - X_{\text{reli}\bullet\bullet}) + \gamma_{23}(\text{Southern, conservative}_j - X_{\text{location}\bullet\bullet}) + \gamma_{23}(\text{Rank}_j - X_{\text{rank}\bullet\bullet}) + u_{2j}$$

$$\beta_{3j} = \gamma_{30}$$

$$\beta_{4j} = \gamma_{40} + \gamma_{41}(\text{Medicine}_j - X_{\text{med}\bullet\bullet}) + \gamma_{42}(\text{Religion}_j - X_{\text{reli}\bullet\bullet}) + \gamma_{43}(\text{Southern, conservative}_j - X_{\text{location}\bullet\bullet}) + \gamma_{43}(\text{Rank}_j - X_{\text{rank}\bullet\bullet}) + u_{4j}$$

The results of the final model for the P scores and the N2 scores are reported in Tables 5 and 6, respectively. The proportions of reduction in the variation in intercept, education slope, and political orientation slope for the P score obtained by modeling the variables for educational contexts are 24.3%, 44.3%, and 11.7%, respectively. Therefore, 24.3% of variation in the average P scores between groups is accounted for by the modeled contextual effects. The results also indicate that the effect of the variable indicating the southern, politically conservative states accounts for a statistically significant variation in the within-group intercepts, β_{0j}s. This suggests that an individual in a higher education institution located in a southern, politically conservative state tends to obtain a P score lower than the overall average for the sample. The specific estimate of this difference is 3.6 points. In addition, the effect of the academic ranking variable is significant, meaning that, on average, the average P score increases 1.5 points as institutional academic ranking increases. For example, this finding indicates that the average P score differs 1.5 points between non–research-oriented PhD-granting institutions and research-oriented PhD-granting universities. Further, the effect of a medical variable on DIT-2 scores is statistically significant, with an associated coefficient indicating a 5.2 point difference. Thus, for DIT-2 data collected from medical schools, it is expected to be on average 5.2 P score points higher than those scores obtained at nonmedical schools. Finally, the result shows that there is a significant unexplained variation remaining in within-group intercepts among groups. This implies that there are other contextual effects that can explain the variation in the institutional averages of P scores not included in the current set of variables.

In addition to the findings that focus on the intercept, HLM provides an index of how the predictors account for variation in slope coefficients. Specifically, 44.3% of the total variation in the magnitude of the relationship between educational level and the P score between groups can be explained by the modeled variables for educational contexts. Among these variables, being in a medical major showed a significant effect for explaining the variation in the education slopes among groups. This indicates that the relationship between educational level and the P score is stronger for people with majors in

Table 5
Summary of HLM Analyses of the Final Model for Variables Predicting the Postconventional Score

	Fixed effect				
	Coeff.	*SE*	*t*	Approx. *df*	*p*
Intercept					
Intercept	37.42	0.78	47.87	60	<.001
Major (medicine)	4.30	2.56	1.68	60	.097
Religious affiliation	3.61	1.87	1.93	60	.058
Southern, conservative	−4.10	1.27	−3.24	60	.002
Academic rank	1.46	0.59	2.46	60	.017
Gender					
Intercept	−4.43	0.39	−11.35	7,268 [a]	<.001
Education					
Intercept	2.04	0.32	6.49	60	<.001
Major (medicine)	1.90	0.58	3.25	60	.002
Religious affiliation	0.14	0.76	0.18	60	.859
Southern, conservative	−0.12	0.50	−0.25	60	.807
Academic rank	−0.20	0.24	−0.87	60	.390
Language status					
Intercept	4.78	1.00	4.75	7,268 [a]	<.001
Political orientation					
Intercept	−1.82	0.29	−6.17	60	<.001
Major (medicine)	−0.67	0.68	−0.99	60	.327
Religious affiliation	0.80	0.87	0.91	60	.366
Southern, conservative	0.62	0.59	1.06	60	.296
Academic rank	−0.45	0.29	−1.58	60	.118

	Random effect			
	SD	χ^2	*df*	*p*
Within-group residual	13.98			
Intercept	5.87	1,238.53	53	<.001
Education	0.70	80.52	53	.009
Political orientation	1.20	107.25	53	<.001

Note. HLM = hierarchical linear modeling; coeff. = coefficient; approx. = approximate.
[a] The estimates of gender and language status were computed based on the number of total individual cases because the relations of gender and language variables with Defining Issues Test 2 scores are constant among institutions. Thus, their associated *df*s were also based on the number of total individual cases.

medicine. In fact, the significant effect of a medical major can be interpreted as the difference between the DIT-2 scores of medical students and other graduate students, because all medical students are graduate students. Thus, the DIT-2 score is found to increase by 2.2 points as the individual education level increases. For example, the expected difference between the DIT-2 scores of senior college students and graduate students is about 2.2 points. Because of the significant cross-sectional interaction effect of the major variables, the amount of increase in the DIT-2 score for medical students should be adjusted upward by 1.9 points in relation to other graduate students. This indicates that there is a variation in moral judgment between graduate programs. Recall that a medical major variable explained the variation in the conditional averages of P scores, and on average, P scores of medical students are about 5.2 points higher than those of nonmedical students. Because of the significant effect of a medical variable on education slope, the expected difference between the DIT-2 scores of medical students and those of other graduate students is about 7.1 points when all other conditions are controlled. Similarly, the expected difference between the DIT-2 scores of senior college students and medical students is found to be 9.3 points. The results also reveal that there is a significant variation remaining for the education slope. This suggests that there exist additional contextual differences that can explain the variation in the magnitude of education slopes among groups.

Finally, although 11.7% of the variation in the magnitude of the relationship between political orientation and the P score is explained by the modeled predictors, none of the modeled contextual variables is significant at $\alpha = .05$. Thus, we should expect that other contextual indicators need to be identified to explain variations in the political orientation slopes. A deviance test for the model fit (Snijders & Bosker, 1999) indicates that the model fit obtained with this model is significantly better than that obtained with the random coefficient model for both the P score, $\chi^2(12) = 28.10$, $p < .01$, and N2 score, $\chi^2(12) = 24.50$, $p = .01$. However, the results of these HLM analyses suggest that the underlying models needs to be expanded to identify additional individual characteristics and educational contexts.

Discussion

It is well known within psychology that individuals interact with and are affected by their contexts. Research in moral psychology is no exception to this overall perspective (Rest, Narvaez, Bebeau, & Thoma, 1999b). With this perspective in mind, the essential goal of the current study was to understand the nature of an individual's moral judgment in the context captured by individual and educational settings. The study posed four questions. In this section, the answer to each of these questions is summarized; then, each finding is discussed.

Table 6
Summary of HLM Analyses of the Final Model for Variables Predicting the N2 Score

	Fixed effect				
	Coeff.	*SE*	*t*	Approx. *df*	*p*
Intercept					
Intercept	36.79	0.84	43.62	60	<.001
Major (medicine)	5.22	2.61	2.00	60	.050
Religious affiliation	3.99	2.01	1.98	60	.052
Southern, conservative	−3.63	1.40	−2.60	60	.012
Academic rank	1.47	0.66	2.24	60	.029
Gender					
Intercept	−4.85	0.41	−11.75	7,262 [a]	<.001
Education					
Intercept	2.16	0.31	7.06	60	<.001
Major (medicine)	1.83	0.64	2.84	60	.007
Religious affiliation	0.18	0.69	0.26	60	.796
Southern, conservative	−0.01	0.53	−0.01	60	.992
Academic rank	−0.16	0.24	−0.68	60	.499
Language status					
Intercept	5.34	0.99	5.42	7,262 [a]	<.001
Political orientation					
Intercept	−1.17	0.28	−4.23	60	<.001
Major (medicine)	−0.44	0.54	−0.82	60	.416
Religious affiliation	0.68	0.83	0.82	60	.415
Southern, conservative	0.67	0.52	1.28	60	.204
Academic rank	−0.31	0.25	−1.24	60	.220

	Random effect			
	SD	χ^2	*df*	*p*
Within-group residual	13.75			
Intercept	6.41	1,446.95	53	<.001
Education	0.83	83.41	53	.005
Political orientation	0.94	96.63	53	<.001

Note. HLM = hierarchical linear modeling; coeff. = coefficient; approx. = approximate.
[a]The estimates of gender and language status were computed based on the number of total individual cases because the relations of gender and language variables with Defining Issues Test 2 scores are constant among institutions. Thus, their associated *df*s were also based on the number of total individual cases.

Does the Average Moral Judgment Level Vary Across Institutions?

The results of the first model indicate that there is a significant variation in the average DIT-2 scores across groups obtained from different higher education institutions. Therefore, as discussed by King and Mayhew (2002), the study supports the view that different types of higher education institutions produce variation in moral judgment development as measured by the DIT-2. However, we found that the amount of variation due to the contextual difference is more modest (i.e., 17.4% of the total variation), relative to the individual differences in moral judgment.

Do DIT-2 Scores Vary According to Demographic Variables, Such As Educational Level, Gender, Political Orientation, and Language Status?

This question was addressed in multiple steps. First, the study examined the existence of relationships between individual characteristics and individual moral judgment. As expected from previous findings (e.g., Bebeau et al., 2003; Rest, Narvaez, Bebeau & Thoma, 1999b), the results of the second model (i.e., the random coefficient model) support the relationship between the modeled individual characteristics and DIT-2 scores. Specifically, the study found a stronger relationship between gender and moral judgment level and between language and moral judgment level than between political orientation and moral judgment or between educational level and moral judgment.

It is interesting that in the second step we found that the strength of the effects of gender and language status is consistent across institutions. This finding suggests that we should expect gender and language effects among individuals to remain relatively constant, regardless of the institutional context. The results indicate that, on average, females tend to obtain P scores and N2 scores that are, respectively, about 4.40 points and 4.83 points higher than those obtained by males. Similarly, native English speakers tend to have P and N2 scores that are approximately 5 points higher than those of nonnative English speakers.

On the other hand, the effects of education and political orientation vary across institutions. Thus, it appears that contextual differences are related to the strength of education and political orientation effects. Finding that contextual difference influences the strength of both education and political orientation effects is consistent with recent views concerning the ways in which the larger social–political context and moral judgment

development interrelate. For instance, regarding educational contexts, it is well known that educational environments vary in terms of the connection between the formal and informal curriculum, which in turn relates to student moral judgment development (e.g., Derryberry & Thoma, 2000; Rest, 1986). Similarly, Narvaez, Getz, Rest, & Thoma (1999) found that the relationship between individual political orientation and moral judgment development is influenced by larger social–political context effects. Specifically, they found that in more politically conservative climates—particularly when religious orthodoxy is also evident—the individual is less likely to question social norms and authority figures. Environments that are open to debate and scrutiny are associated with growth in moral judgment development (e.g., Thoma, 2006; Narvaez et al., 1999). Overall, therefore, it seems plausible that the institutions assessed by our study vary on within-institutional factors and are differentially influenced by broader social–political environments; thus, it is not surprising that we observed variation in the strength of the relationship between the educational levels, political orientation, and moral judgment development across institutions.

If There Is Variation in the Average Moral Judgment Level Across Educational Contexts, To What Extent Is the Larger Context Related to an Individual's DIT-2 Scores?

Finding that there is significant variation in institutional average P scores, the analyses then addressed whether the obtained variation is related to the location of the higher education institution where DIT data were collected. Consistent with Rest's (1979a) finding using DIT-1 with students in the 1970s and 1980s, our results also indicate that the southern states are associated with lower DIT-2 scores. Unlike Rest's (1979a) earlier work, which included only two southern states (Georgia and Virginia), the current study obtained data from 28 different states, including 9 southern states (i.e., Alabama, Florida, Georgia, Kentucky, Maryland, Tennessee, Texas, Virginia, and West Virginia). Because the current southern sample is more extensive than the Rest sample, the results of this study have greater generalizability.

In addition, students in medical schools are associated with higher average moral judgment levels when compared to students in business and other majors. Possibly the fact that such schools are preparing students for a profession or calling that is essentially a moral enterprise serves as an environmental factor that influences level of moral judgment. Alternatively, it may be that a medical institution's use of a measure of moral judgment to assess its students indicates an institutional commitment to moral education that is reflected in student responses. Further, students at institutions with higher academic ranking tend to show, on average, higher moral judgment level than those at schools with lower ranking. Although an explanation for this finding is speculative, more highly ranked schools often assume a leadership position in medical education (Beauchamp & Childress, 2001). Given the current interest in medical ethics within the health profession, these flagship schools may be more likely to incorporate a well-articulated ethics component in the curriculum and to have had it in place for a longer time.

The average DIT-2 score for religiously affiliated institutions was not significantly lower than non–religiously affiliated institutions. On the contrary, the descriptive statistics indicate that the average DIT-2 scores for religiously affiliated institutions are higher, though not significantly so, than those for non–religiously affiliated institutions. This finding is consistent with studies by McNeel (1994) and Pascarella and Terenzini (1991, 2005), who argued that whether an institution enhances moral judgment development depends not on whether the institution has a religious orientation, but on whether the college genuinely focuses on a liberal arts education.

If There Is Variation in the Strength of the Relationship Between DIT-2 Scores and a Particular Individual Demographic Characteristic Across Educational Contexts, to What Extent Does the Contextual Difference Explain This Variation?

The medical school setting is the only major variable that accounts for variation across contexts. Specifically, medical schools tend to show a stronger relationship between education and moral judgment than other graduate programs. It is interesting that during the time covered by this study, medical ethics coursework has become a mainstay of medical education (Bebeau & Monson, 2008). Thus, the effect noted in the current study may be due to the targeted coursework in medical ethics. It is also possible that the obtained effect is due to the types of students selected into these degree programs. As a next step, it would be interesting to investigate whether student selection effects can account for the findings and, in so doing, tease out whether a setting or ability explanation best fits the data.

Limitations

It is important to note that the information regarding educational contexts was rather limited. Therefore, although the unexplained variations in intercept (i.e., conditional average DIT scores), education, and political orientation still remain, it was impossible to model additional variables to account for the remaining variation. As noted earlier, the additional unexplained variation is not trivial. Therefore, further investigations designed to identify context effects on individual moral judgment development should extend and elaborate the way in which educational environments are measured. The current study helps to support such efforts and provides an analysis strategy to investigate systematically the relationship between educational contexts and individual moral judgment development.

It should also be noted that possibly influential variables, such as individual academic achievement indicated by college GPA or standardized achievement score, were not included in the analysis because this information was not collected as part of the archived data sets. The lack of ability information is problematic because of the known relationship between ability measures and moral judgment development, as well as the clear link between ability and selection into various educational contexts. More specifically, the validity of the reported relationship between the investigated demographic characteristics and moral judgment development would decrease when both conditions are met: (a) The variables that are not taken into account (e.g., GPA) are highly related to the DIT-2 P score, and (b) the variables that are not in the analysis are highly correlated

with the variables that are in question (e.g., Mauro, 1990). For example, according King and Mayhew (2004), the correlation between academic achievement and DIT scores reported in seven articles is weak to moderate. Although it is clear that there is an increase in academic ability during college years (Pascarella & Terenzini, 2005), the correlation between academic achievement and gender, political orientation, and language status is less well known. Current evidence suggests that the magnitude of the relationship tends to be small (Rest, 1979a, 1986). Taken together, we acknowledge that these unmeasured variables and self-selection into specific educational contexts most likely have an effect on the estimates of the relationships between individual demographic characteristics and moral judgment. However, we note that our findings are fairly robust, and thus it is unlikely that ability can account for the patterns we do observe. Indeed, our study should encourage others to extend this work and account for other important variables that relate to moral judgment level.

Conclusion

In 2002, King and Mayhew wrote, "Given the complexity of both the [moral judgment] construct and of institutional contexts, larger-scale studies utilizing more sophisticated statistical techniques are needed to untangle factors that lead to the development of moral judgment" (p. 265). We concur. This study is the first attempt to understand how individual characteristics and educational contexts together relate to moral judgment development. Our findings contribute to at least two aspects of moral judgment research. First, the findings empirically demonstrate that the educational context should be taken into account to understand variation in the individual level of moral judgment. Thus, it seems clear that by ignoring the context in which the data were collected we run the risk of a biased, incomplete picture of individual moral judgment development. As such, future models of moral judgment development in late adolescence and youth must take into account a more complex understanding of context to explain fully individual development. As described here, characteristics of the specific institution as well as the larger social–political climate are associated with variation in individual scores, and both must be considered to account for an appropriate model of context.

By contrast, our results also support the consistency of an effect on moral judgment development across contexts due to individual characteristics like gender and language. That is, gender and primary language do not vary by the context effects mentioned above; thus, they should become a part of our expectations when collecting data in populations of individuals in their late teens through their 30s. Indeed, given the strengths of our findings, if DIT-2 users do not find the expected gender difference, they may conclude that the characteristics of their respondents are uniquely different from students in the general population of institutions of higher education. Therefore, one important aspect of our findings is to give future DIT-2 users an indication of what to expect when interpreting the relationship between the most common demographic categories and DIT-2 scores. Indeed, this aspect of our work can be interpreted as analogous to a meta-analysis and can be used to estimate effects in the population. Taken together, we hope the results of these findings alert researchers and practitioners to the important effects of educational environments, giving environmental effects serious consideration when interpreting moral judgment development in higher education.

References

Beauchamp, T. L., & Childress, J. F. (Eds.). (2001). *Principles of biomedical ethics.* Oxford, England: Oxford University Press.

Bebeau, M. J. (2002). The Defining Issues Test and the four component model: Contributions to professional education. *Journal of Moral Education, 31,* 271–295.

Bebeau, M. J., Maeda, Y., & Tichy-Reese, M. (2003, March). *DIT-2: Moral schema norms by level of education, gender and political orientation.* Paper presented at the annual meeting of the Association for Practical and Professional Ethics, Charlotte, NC.

Bebeau, M. J., & Monson, V. E. (2008). Guided by theory, grounded in evidence: A way forward for professional ethics education. In L. Nucci & D. Narvaez (Eds.), *Handbook of moral and character education* (pp. 557–582). New York: Routledge.

Bebeau, M. J., & Thoma, S. J. (2003). *Guide for DIT-2: A guide for using the Defining Issues Test, Version 2 (DIT-2) and the scoring service of the Center for the Study of Ethical Development.* Minneapolis, MN: Center for the Study of Ethical Development.

Carnegie Foundation for the Advancement of Teaching. *The Carnegie classification of institutions of higher education.* Retrieved November, 8, 2007, from http://www.carnegiefoundation.org/classifications/index.asp

Crowson, H. M., DeBacker, T., & Thoma, S. J. (2007). Are DIT scores empirically distinct from measures of political identification and intellectual ability? A test using post-9/11 data. *British Journal of Developmental Psychology, 25,* 197–211.

Deemer, D. K. (1989). Moral judgment and life experience. *Moral Education Forum, 14,* 11–21.

Derryberry, P., & Thoma, S. (2000). The friendship effect: Its role in the development of moral thinking in students. *About Campus, 5,* 13–20.

Emler, N., Resnick, S., & Malone, B. (1983). The relationship between moral reasoning and political orientation. *Journal of Personality and Social Psychology, 45,* 1073–1080.

Good, J. L., & Cartwright, C. (1998). Development of moral judgment among undergraduates. *College Student Journal, 32,* 270–277.

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

King, P. M., & Mayhew, M. J. (2004). Theory and research on the development of moral reasoning among college students. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research: Vol. 19* (pp. 375–440). Edison, NJ: Agathon Press.

Landsman, M., & McNeel, S. P. (2004). Moral judgment of law students across three years: Influences of gender, political ideology and interest in altruistic law practice. *South Texas Law Review, 45,* 891–920.

Marx, B. (2005). *Understanding the relationship of moral reasoning and liberalism–conservatism.* Unpublished doctoral dissertation, University of Sydney, Australia.

Mauro, R. (1990). Understanding L.O.V.E. (left out variable error): A method for estimating the effect of omitted variables. *Psychological Bulletin, 108,* 314–329.

McNeel, S. P. (1991). Christian liberal arts education and growth in moral judgment. *Journal of Psychology and Christianity, 10,* 311–322.

McNeel, S. P. (1994). College teaching and student moral development. In J. Rest & D. Narvaez (Eds.), *Moral development in the professions: Psychology and applied ethics* (pp. 27–50). Hillsdale, NJ: Erlbaum.

Narvaez, D., Getz, I., Rest, J., & Thoma, S. J. (1999). Orthodoxy, progressivism, and development of moral judgment. *Developmental Psychology, 35,* 478–488.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects student:*

Findings and insights from twenty years of research. San Francisco: Jossey-Bass.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research.* San Francisco: Jossey-Bass.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rest, J. R. (1979a). *Development in judging moral issues.* Minneapolis, MN: University of Minnesota Press.

Rest, J. R. (1979b). *Defining Issues Test.* Minneapolis, MN: University of Minnesota.

Rest, J. R. (1986). *Moral development: Advances in research and theory.* New York: Praeger.

Rest, J. R., & Narvaez, D. (1998). *Defining Issues Test 2.* Minneapolis, MN: University of Minnesota.

Rest, J., Narvaez, D., Bebeau, M., & Thoma, S. J. (1999a). A neo-Kohlbergian approach: The DIT and schema theory. *Educational Psychological Review, 11,* 291–324.

Rest, J. R., Narvaez, D., Bebeau, M. J., & Thoma, S. J. (1999b). *Postconventional moral thinking: A neo-Kohlbergian approach.* Mahwah, NJ: Erlbaum.

Rest, J., Narvaez, D., Thoma, S. J., & Bebeau, M. (1999). DIT-2: Devising and testing a revised instrument for moral judgment. *Journal of Educational Psychology, 91,* 644–659.

Rest, J. R., & Thoma, S. J. (1985). The relation of moral judgment development to formal education. *Developmental Psychology, 21,* 709–714.

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

Self, D. J., & Baldwin, D. C., Jr. (1998). Does medical education inhibit the development of moral reasoning in medical students? A cross-sectional study. *American Medicine, 73,* 91–93.

Self, D. J., Pierce, A., & Shadduck, J. A. (1995). Description and evaluation of a course in veterinary ethics. *Journal of the American Veterinary Medical Association, 207,* 1550–1552.

Snijders, T. A. B., & Bosker, R. J. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* London: Sage.

Thoma, S. J. (1986). Estimating gender differences in the comprehension and preference of moral issues. *Developmental Review, 6,* 165–180.

Thoma, S. J. (2006). Research on the Defining Issues Test. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 67–92). Mahwah, NJ: Erlbaum.

Thoma, S., Barnett, R., Rest, J., & Narvaez, D. (1999). What does the DIT measure? *British Journal of Social Psychology, 38,* 103–111.

Thoma, S. J., & Ladewig, B. H. (1993, April). *Close friendships, friendship networks, and moral judgment development during the college years.* Paper presented to the Society for Research in Child Development, New Orleans, LA.

Walker, L. J. (2006). Gender and morality. In M. Killen & J. Smetana (Eds.), *Handbook of moral development* (pp. 93–118). Mahwah, NJ: Erlbaum.

White, R. D., Jr. (1999). Are women more ethical? Recent findings on the effects of gender upon moral development. *Journal of Public Administration Research and Theory, 9,* 459–471.

Wilson, F. (1995). The effect of age, gender and ethnic/cultural background on moral reasoning. *Journal of Social Behavior and Personality, 10,* 67–78.

Appendix A

The small village in northern India has experienced shortages of food before, but this year's famine is worse than ever. Some families are even trying to feed themselves by making soup from tree bark. Mustaq Singh's family is near starvation. He has heard that a rich man in his village has supplies of food stored away and is hoarding food, while its price goes higher, so that he can sell the food later at a huge profit. Mustaq is desperate and thinks about stealing some food from the rich man's warehouse. The small amount of food that he needs for his family probably wouldn't even be missed.

Appendix B

The Random Effects ANOVA

Within-group model: $Y_{ij} = \beta_{0j} + r_{ij}$

Between-groups model: $\beta_{0j} = \gamma_{00} + u_{0j}$

The within-institution model involves three terms. Y_{ij} is the DIT-2 P score (or N2 score) of the *i*th individual in the *j*th institution; β_{0j} is the within-group random coefficient, which represents the average P score or N2 score (intercept) for the *j*th institution; r_{ij} is an individual-level residual. Thus, each institution has a unique regression equation to predict the P score for the *i*th individual in the *j*th institution.

The between-groups model also consists of three terms. The within-group random coefficient, β_{0j}, now serves as the dependent variable; γ_{00} is the fixed coefficient, which is the overall average P score (or N2 score) across institutions; u_{0j} represents the residual of the *j*th institution.

The Random Coefficient Model

Within-group model:

$$Y_{ij} = \beta_{0j} + \beta_{1j}\,(\text{Gender}_{ij} - X_{\text{Gender}\bullet j}) + \beta_{0j}\,(\text{Political Orientation}_{ij} - X_{\text{Politics}\bullet j}) + \beta_{3j}\,(\text{Language Status}_{ij} - X_{\text{Language}\bullet j}) + \beta_{4j}\,(\text{Education Level}_{ij} - X_{\text{Education}\bullet j}) + r_{ij}$$

Between-groups model:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}$$

$$\beta_{2j} = \gamma_{20} + u_{2j}$$

$$\beta_{3j} = \gamma_{30} + u_{3j}$$

$$\beta_{4j} = \gamma_{40} + u_{4j}$$

Because the individual demographic variables were centered around the group mean, β_{0j} represents the group average of DIT-2 scores. β_{1j} is the slope capturing the effect of gender on the DIT-2 scores of individuals within the *j*th group, holding other variables in the model constant. In other words, β_{1j} represents the unique contribution of gender difference to the DIT-2 scores when the contribution of all other modeled predictors is statistically controlled. Thus, the coefficient can be interpreted as the difference between the DIT-2 scores of a female student and a male student who have exactly the same characteristics with regard to educational level, political orientation, and language status. $\bar{X}_{1j}$ is the average of the gender variable in the *j*th group. Because the gender variable is a dichotomous variable, $\bar{X}_{1j}$ is the proportion of females in the *j*th group. The remaining variables in the model are interpreted in the same way.

Received April 24, 2007
Revision received July 2, 2008
Accepted July 7, 2008 ■