Res High Educ (2012) 53:26–46
DOI 10.1007/s11162-011-9226-3

Going Deep into Mechanisms for Moral Reasoning Growth: How Deep Learning Approaches Affect Moral Reasoning Development for First-year Students

Matthew J. Mayhew · Tricia A. Seifert · Ernest T. Pascarella · Thomas F. Nelson Laird · Charles F. Blaich

Received: 7 July 2010 / Published online: 2 April 2011

Abstract The purpose of this paper was to determine the effects of deep approaches to learning on the moral reasoning development of 1,457 first-year students across 19 institutions. Results showed a modest positive relationship between our measures of deep approaches to learning and moral reasoning at the end of the first year of college even after controlling for precollege moral reasoning. After accounting for a host of demographic and relevant student characteristics and for the natural clustering of students, we found that the integrative learning subscale, which captures students' participation in activities designed to integrate information from varied sources and diverse perspectives, positively affected moral reasoning among first-year students. Implications for researchers and practitioners are discussed.

Keywords Moral reasoning development · Deep learning · Quantitative · First-year students

Despite the ubiquity of findings linking formal participation in college to moral reasoning development (see Burwell et al. 1992; Foster and LaForce 1999; Mentkowski and Strait 1983; Rest 1979a; Shaver 1985, 1987; Whiteley 1982), very little research has explored

M. J. Mayhew (✉)
New York University, 239 Greene Street, Suite 300, New York, NY 10012, USA
e-mail: matt.mayhew@nyu.edu

T. A. Seifert
Ontario Institute for Studies in Education, University of Toronto, Toronto, ON M5S 1V5, Canada

E. T. Pascarella
University of Iowa, Iowa City, IA 52242, USA

T. F. Nelson Laird
Indiana University at Bloomington, Bloomington, IN 47405-3905, USA

C. F. Blaich
Center of Inquiry in the Liberal Arts at Wabash College, Crawfordsville, IN 47933, USA

educational processes, like deep approaches to learning, as likely determinants of moral reasoning development. This is somewhat surprising given calls by student development scholars and practitioners for more research focusing on how development occurs, not just that it does. As King (2005) noted, "Clearly, there is a need for focused research attention on questions of the process as well as the content of student development (p. 49)." The purpose of this study is to examine the impact of one such mechanism, deep approaches to learning, on moral reasoning development.

Moral reasoning is "a psychological construct that characterizes the process by which people determine that one course of action in a particular situation is morally right and another course of action is wrong" (Rest et al. 1997, p. 5). Structurally, the moral reasoning construct or what neo-Kohlbergians (see Rest et al. 1999) have called "moral reasoning schema" reflects an individual's cognitive representation of justice that is susceptible to change or re-organization expressed by the individual's ability to construct and reconstruct new moral responses based on challenging and often unfamiliar environmental cues (Kohlberg 1976). As moral reasoning develops, frames for understanding justice expand from systems of fairness that serves oneself to ones that serve known others to ones that serve anonymous others (see Kohlberg 1976).

In existing theories of moral reasoning development, Kohlberg (1976) and his contemporaries (e.g., Rest et al. 1979a, b) were keenly interested in the mechanisms needed to promote moral reasoning development. Charging researchers with the task of understanding the specific contextual factors comprising any given intervention, Kohlberg (1976) and Rest (1979a, b, 1988) underscored the importance of learning experiences that disrupted existing cognitive structures for understanding justice by introducing new, often more complex, information that could not be easily assimilated into familiar thought patterns.

The cognitive work induced by these disruptions is similar to that required by a deep approach to learning. Deep approaches to learning create spaces for students to explore the unfamiliar, often engendered by exposure to diverse perspectives, and to subsequently synthesize and integrate this information into new frameworks for making meaning of the self as well as external stimuli (Ramsden 2003; Tagg 2003). As Tagg (2003) articulated, "Deep learning is learning that takes root in our apparatus of understanding, in the embedded meanings that define us and that we use to define the world" (p. 70).

Faculty who encourage deep approaches to learning by their students, often through teaching and assessment practices that foster active participation (including interacting with peers) over an extended period, ask students to routinely engage divergent perspectives for eventual integration into the frames these students use to understand the world and themselves (Biggs 2003; Ramsden 2003). In addition, faculty with pedagogies reflecting deep learning approaches help students engage learning deeply by shifting from instructor-dominated pedagogies to active, learner-centered activities that can take students to more complex understandings and new meanings as they apply what they learn to real life examples (Lave and Wenger 1991; Tagg 2003).

As a result of these faculty practices, students affected by deep approaches to learning often express a personal commitment to challenge one's own understanding of the world by engaging the unfamiliar. Students demonstrate such a commitment by reading widely, integrating information across resources and contexts, discussing ideas with others, reflecting on how individual pieces of information relate to larger constructs or patterns, and applying knowledge in real world situations (Biggs 1987, 1989, 2003; Entwistle 1981; Ramsden 2003; Tagg 2003).

What makes the current study unique is its examination of deep approaches to learning as mechanisms for moral reasoning development. Theoretically, we posit that neither

moral reasoning development nor deep approaches to learning can occur in the absence of unfamiliar, external stimuli forcing students to re-examine what they know. Second, we contend that deep approaches to learning, with their imbedded engagement with concepts that challenge the learner's current understandings, including how students make meaning of justice and fairness, may also be related to, if not spur, moral reasoning development. It is our hope that student development theorists and educators will find some utility to this study in its attempts to explain why moral reasoning occurs, not just that it does.

Literature Review

We adopted a multi-phased approach to reviewing the literature used to empirically support this study. First, we deconstructed deep approaches to learning into its three component parts: higher-order, integrative, and reflective learning (Nelson Laird et al. 2006, 2008). Using these components as our guide, we reviewed the literature on moral reasoning development as it related to each component. Next, we reviewed the moral reasoning literature and that associated with deep approaches to learning to justify the variables included in the conceptual framework guiding this study.

As a result of this multi-phased process of reviewing the literature, we argue that the literature review served three purposes: to conceptually link components of deep learning approaches to those related to moral reasoning development; to empirically justify the variables included in the conceptual and subsequent operational frameworks; and to provide a rationale for analytic decisions, including testing for conditional effects, adopted for this study. We turn now to this literature review.

Connecting Components of Deep Approaches to Learning to Moral Reasoning

In recent work connected to the National Survey of Student Engagement (NSSE 2006), deep approaches to learning were deconstructed into three constituent elements: practices that spur higher-order, integrative, and reflective learning (Nelson Laird et al. 2006, 2008). Though not using the NSSE measures, a series of studies have investigated variations of these elements and their respective effects on moral reasoning development. Interestingly, these studies have approached learning from a series of perspectives, with some using cognitive outcomes as proxies for learning and others investigating educational practices related to the deep approaches to learning sub-components, like reflective or integrative learning as examples. We summarize the findings from these studies below.

That a certain level of cognitive sophistication is a necessary but insufficient condition for moral reasoning growth was a finding common to many studies linking cognitive to moral development. What constitutes cognitive development differs among scholars; however, King and Mayhew (2004) offered a rubric for deconstructing cognitive development in relation to moral reasoning, using "three distinct types" (p. 405): personal epistemology, academic achievement and logical reasoning. Across these studies, findings overwhelmingly supported the notion that high levels of moral reasoning development were associated with more complex, higher-order expressions of cognition.

Practices that challenge students to be more reflective have also been linked to moral reasoning development. In his seminal work entitled, *Ethics*, Dewey (1933) concisely stated, "It [reflection] is a means to essentially moral ends" (p. 124). Building upon this idea, Rest (1987, 1988) and Rest and Deemer (1986) asked students to identify what they thought had influenced their moral reasoning; participants who demonstrated greater moral

reasoning gains cited "spending more time contemplating issues" more frequently than peers exhibiting fewer gains, suggesting that reflection was a key factor in promoting moral reasoning. More recently, in their study of educational practices linked to moral reasoning development, Mayhew and King (2008) found that, when compared to peers who reported fewer opportunities for reflection, students enrolled in classes with faculty who created more of these opportunities were more likely to experience developmental gains in moral reasoning.

Less is known about integrative learning and its effects on moral reasoning development. Although not directly measured, one study implied that integrated learning experiences may, in fact, be related to moral reasoning growth. A distinctive learning environment, Alverno College offers an integrated, ability-based curriculum evaluated for its impact on deep and sustained college learning; students exposed to this curriculum demonstrated developmental gains in moral reasoning (Mentkowski et al. 2000; Mentkowski and Strait 1983; Rogers 2002). Interviews with students corroborated the findings linking the integrated curriculum with moral reasoning gains and drove Mentkowski et al. (2000) to conclude that students "came to appreciate and understand differing values because they were repeatedly asked to examine and discuss them" across learning contexts (p. 130).

Variable Justification

We reviewed over two hundred studies for justifying the variables included in the conceptual and operational models guiding this study. Rather than provide a full review of each study (see King and Mayhew 2002, 2004 for such a review), we present examples of research that highlight the key findings used to justify variables selected for inclusion in this study. Consistent with other studies examining deep approaches to learning (Nelson Laird et al. 2008), we organized these examples into sections according to Biggs (2003), 3-P model (presage, process, and product).

Presage Covariates

A series of presage covariates were explored for their potential in explaining moral reasoning development and deep approaches to learning. From the moral reasoning literature, it appears as though gender, political orientation, precollege ACT score or equivalent, precollege interest in cognitively complex activities, and precollege academic motivation shared significant relationships with moral reasoning development.

Interestingly, empirical efforts designed to link these presage covariates with moral reasoning development were largely validation studies, performed in response to claims that Kohlbergian-based theory or subsequent measurement were biased. For example, perhaps the most scholarly attention toward understanding precollege effects on moral reasoning emerged from gender studies in response to Gilligan's (1977) initial claims of gender bias in Kohlberg's theory of moral reasoning development. Research has since shown that women consistently outscore men on moral reasoning measures (see Brabeck 1983; King and Mayhew 2002, 2004; Walker 1984); such a finding has led scholars to dismiss claims of gender bias in the measure of moral reasoning (Walker 2006). Despite recent dismissals of gender-bias claims, we included gender in our conceptual and operational models due to the overwhelming evidence suggesting that women have higher moral reasoning scores than men.

Some scholars have expressed concerns that the measure of moral reasoning was a proxy measure of political orientation or verbal ability. Studies of the former have shown that moral reasoning is indeed a construct distinctive from political orientation. However, individuals identifying as politically liberal have consistently outscored those identifying as politically conservative (Barnett et al. 1995; Elmer et al. 1983, 1997; Fisher and Sweeney 1998; Murk and Addleman 1992; Narvaez et al. 1999; Thoma 1993). For this reason, we included precollege political orientation as a presage covariate.

Turning to verbal ability, a series of studies have demonstrated that although conceptually distinct, moral reasoning and verbal ability are related, with individuals scoring higher on measures of verbal ability more likely to score higher on moral reasoning measures (Gongre 1981; Hendel 1991; Quarry 1997; Sanders et al. 1995; Stepp 2002). A possible outgrowth of scholars attempting to understand the nuanced relationship between measures of verbal ability and moral reasoning, a number of studies have emerged to examine dimensions of cognition and their relationship to moral reasoning. An example of one such dimension includes cognitive and academic motivation (Crowson 2004; King and Mayhew 2002, 2004; Mayhew and King 2008; Mayhew et al. 2008, 2010, in press), with individuals highly motivated to perform well on cognitive and academic tasks demonstrating higher moral reasoning scores than less motivated individuals. To isolate the effects of deep learning approaches on moral reasoning development, we included three cognitive-related measures in our study: precollege ACT score or equivalent, precollege interest in cognitively complex activities, and precollege academic motivation.

Although race and socioeconomic status have consistently been shown not to share significant relationships with moral reasoning development (see King and Mayhew 2004); both race and parents' education, an indicator of socioeconomic status, were empirically linked to deep approaches to learning in the Nelson Laird et al. (2008) models. Therefore, these variables were included in our conceptual and operational frameworks as controls used to better isolate the distinctive effects of deep approaches to learning on moral reasoning development.

Process Covariates

Course-taking patterns have been shown to have effects on deep approaches to learning (Nelson Laird et al. 2008) and to moral reasoning development (Hurtado et al. 2003; Mayhew and King 2008; Mentkowski et al. 2000; Nevin and McNeel 1992). Since discipline affects faculty emphasis on deep approaches to learning and senior use of these approaches (Nelson Laird et al. 2008), it was important to include indicators of discipline-specific course-taking patterns in the conceptual and analytic frameworks.

Empirical Basis for Analytic Decisions

Due to the theoretical links among presage variables, deep approaches to learning, and moral reasoning development, a series of strategies was implemented to explain this relationship. First, the presage and process variables were included in analytic models as controls, helping to isolate the amount of variance explained in moral reasoning by deep learning approaches. As an additional analytic step, the relationship between deep approaches to learning and moral reasoning development was tested to determine if effects were general (i.e., the same for all students in the sample) or conditioned based on self-identified student characteristics. Not only is such a strategy warranted given theoretical links between presage and process variables (e.g., race and deep learning approaches)

established in the literature but this analytic step is consistent with the college impact heuristic; as Pascarella (2006) argues: "the same intervention or experience might not have the same impact for all students, but rather might differ in the magnitude or even the direction of its impact for students with different characteristics or traits" (p. 512).

Purpose

The conceptual links between moral reasoning development and deep approaches to learning as well as the existence of evidence connecting the sub-components of deep learning approaches to moral reasoning served as the theoretical and empirical basis for this paper. Our purpose was to explore the effects[1] of an overall measure of deep approaches to learning as well as measures of higher-order, integrative, and reflective learning on the moral reasoning development of 1,457 students enrolled across 19 institutions. We administered a battery of instruments to students at the beginning and end of their first year in college and used the resulting data to construct analytic models designed to test the effects of deep approaches to learning on moral reasoning development. Additionally, we examined the extent to which the effect of deep learning approaches on moral reasoning development differed based on students' background characteristics and course-taking patterns.

Methods

Research Questions

One central research question and two auxiliary questions guide this study. They read as follows:

1. How do approaches to deep learning influence moral reasoning development among first-year students?
    a. To what extent does the effect of deep learning approaches on moral reasoning development differ based on students' background characteristics?
    b. To what extent does the effect of deep learning approaches on moral reasoning development differ based on course-taking patterns?

Conceptual Framework

Biggs (2003), in his 3-P model (presage, process, and product), suggests that Student Factors and the Teaching Context lead to students' choice of Learning-Focused Activities (e.g., deep approaches to learning), which in turn affects the Learning Outcomes. This model would encourage our inclusion of student background characteristics (e.g., gender, race, parental education, academic ability), learning environment characteristics (e.g., courses taken), and learning process indicators such as deep approaches to learning. While the Biggs model is not substantially different from other models of college impact (e.g.

[1] We use the term "effect" not in the causal sense of the word but in the statistical sense. The pretest–posttest longitudinal design and our use of statistical controls allowed us to identify that part of the variance in y that we could attribute to x (Shadish et al. 2002).

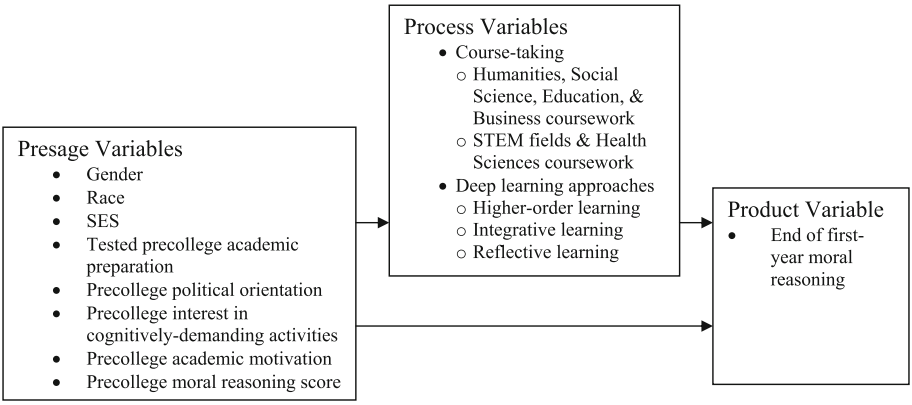

Fig. 1 Conceptual framework describing the relationship between deep learning approaches and moral reasoning development

Astin 1993; Pascarella 1985), it highlights that aspects of students' background and the college context influence students' choices of learning approaches, something central to our study that is not emphasized in the other models.

Our presage variables included a host of covariates. These were: race, gender, parental education, precollege ACT score or equivalent, precollege political orientation, precollege interest in cognitively complex activities precollege academic motivation, and precollege moral reasoning score.

Turning to process variables, we included two measures of course-taking patterns. The course-taking variables measured (a) the number of courses taken in humanities, the social sciences, education, and business; and (b) the number of courses completed in the natural sciences, health, math, and engineering. This parsing allowed us to capture students' exposure to both hard and soft fields, the major source of disciplinary difference found by Nelson Laird et al. (2008) in faculty emphasis on and senior use of deep approaches to learning. For a sample of first-year students, like ours, the measures of course-taking were preferred over capturing major, because many students in the sample had yet to declare a major. In addition, with our focus on first-year students, trying to capture finer distinctions in course taking patterns was difficult without creating highly skewed variables (i.e., many students take zero courses in certain areas). Our other process variables were the variables measuring deep learning approaches.

Our product variable was reported moral reasoning at the end of students' first year in college. See Fig. 1 for the conceptual diagram describing this study.

Samples

Institutional Sample

The sample in the study consisted of incoming first-year students at 19 four-year and two-year colleges and universities located in 11 different states from four general regions of the United States: Northeast, Southeast, Midwest, and Pacific Coast. Institutions were selected from more than 60 colleges and universities responding to a national invitation to participate in the Wabash National Study of Liberal Arts Education (the Wabash Study). Funded by the Center of Inquiry in the Liberal Arts at Wabash College, the Wabash Study

is a large, longitudinal investigation of the effects of liberal arts colleges and liberal arts experiences on the cognitive and personal outcomes theoretically associated with a liberal arts education. The institutions were selected to represent differences in college and universities nationwide on a variety of characteristics including institutional type and control, size, location, and patterns of student residence. However, because the study was primarily concerned with the impacts of liberal arts colleges and liberal arts experiences, liberal arts colleges were purposefully over-represented.

Our selection technique produced a sample with a wide range of academic selectivity, from some of the most selective institutions in the country to some that were essentially open admissions. There was also substantial variability in undergraduate enrollment, from institutions with entering classes between 3,000 and 6,000, to institutions with entering classes between 250 and 500. According to the 2007 Carnegie Classification of Institutions, three of the participating institutions were considered research universities, three were regional universities that did not grant the doctorate, two were 2-year community colleges, and 11 were liberal arts colleges.

Student Sample

The individuals in the sample were first-year, full-time undergraduate students participating in the Wabash Study at one of the 19 institutions in the study. The initial sample was selected in either of two ways. First, for larger institutions, it was selected randomly from the incoming first-year class at each institution. The only exception to this was at the largest participating institution in the study, where the sample was selected randomly from the incoming class in the College of Arts and Sciences. Second, for a number of the smallest institutions in the study—all liberal arts colleges—the sample was the entire incoming first-year class. The students in the sample were invited to participate in a national longitudinal study examining how a college education affects students, with the goal of improving the undergraduate experience. They were informed that they would receive a monetary stipend for their participation in each data collection, and were also assured any information they provided would be kept confidential and would never become part of their institutional records.

Data Collection

Initial Data Collection

The initial data collection was conducted in the early fall of 2006 with 4,501 students from the 19 institutions. This first data collection lasted between 90 and 100 min and students were paid a stipend of $50 each for their participation. The data collected included a Wabash Study precollege survey that sought information on student demographic characteristics, family background, high school experiences, political orientation, educational degree plans, and the like. Students also completed a series of instruments that measured dimensions of cognitive and personal development theoretically associated with a liberal arts education. One of these was a revised version of the Defining Issues Test (DIT2), which measures moral reasoning or judgment (Rest et al. 1999), and is described in greater detail in the "Dependent Variable" section below. Because it is a rather lengthy instrument to complete, and because we were concerned about the use of student time during the assessment, the DIT2 was not completed by all students in the sample. Rather, at each institution it was randomly assigned to half of the student study participants, while the

other random half of the sample completed a different instrument of almost identical length.

Follow-up Data Collection

The follow-up data collection was conducted in spring 2007. This data collection took about two hours and participating students were paid an additional stipend of $50 each. Two types of data were collected. The first was based on questionnaire instruments that collected extensive information on students' experiences in college. Two complementary instruments were used: the NSSE, previously described, and the Wabash Student Experiences Survey (WSES). However, for the purposes of this study, we focus on information provided by the NSSE. The second type of data collected consisted of follow-up (or posttest) measures of the instruments measuring dimensions of cognitive and personal development that were first completed in the initial data collection. All students completed the NSSE and WSES prior to completing the follow-up instruments assessing cognitive and personal development. Both the initial and follow-up data collections were administered and conducted by ACT (formerly the American College Testing Program).

Of the original sample of 4,501 students who participated in the fall 2006 testing, 3,081 participated in the spring 2007 follow-up data collection, for a response rate of 68.5%. These 3,081 students represented 16.2% of the total population of incoming first-year students at the 19 participating institutions.[2] The response rate for returning to complete the DIT2 was 69.7%. To make the sample more reflective of the population from which it was drawn, we created an algorithm that weighted the sample to best approximate the first-year undergraduate student population at each institution. Using information provided by each institution on sex, race, and ACT score or equivalent, follow-up participants were weighted up to each institution's first-year undergraduate population by sex (male or female), race (Caucasian, African American/Black, Hispanic/Latino, Asian/Pacific Islander, or other), and ACT score(or equivalent) quartile. Weights were used in all of our analyses.

Dependent Variable

The dependent variable in the study was an index of moral reasoning derived from student responses to the Defining Issues Test 2 (DIT2). The DIT2 is a revised version of James Rest's original DIT2 from 1979 that measures one component of moral development, known as moral reasoning (Rest et al. 1999). The DIT2 presents several dilemmas about social problems, such as should a starving man steal food for his family from someone who is hoarding resources. After each presented dilemma, respondents are asked to rate and rank a series of 12 reasoning strategies in terms of their importance in making a decision about the scenario. This process of rating and ranking is repeated for five scenarios. Responses are then used to generate an index score, called "N2," that reflects the extent to which one reasons about dilemmas from a system of fairness that serves oneself to one that serves anonymous others. The internal consistency for the N2-score ranges from .77 to .81 (Rest et al. 1999). Scoring of the instrument is completed by the Center for the Study of Ethical Development at the University of Minnesota. Of the 1,493 students who reliably

[2] This is a conservative estimate as it reflects the response rate if every person who was invited to participate in the study made a conscious decision to do so or not. We estimate that up to half of the students in the sample may not have received the invitation letter.

completed the DIT2 in both the fall 2006 and the spring 2007 data collections, useable data for all analyses of DIT2-N2 scores were available for 1,457 students.[3]

Independent Variables

Derived from NSSE by Nelson Laird and colleagues, the independent variables in the study were an overall deep approaches to learning scale and its three subscales (Nelson Laird et al. 2006, 2008). According to Nelson Laird et al. (2008), the four-item Higher-Order Learning subscale "focuses on the amount students believe that their courses emphasize advanced thinking skills such as analyzing the basic elements of an idea, experience, or theory and synthesizing ideas, information, or experiences into new, more complex interpretations" (p. 477). The Integrative Learning Scale consists of five items and measures "the amount students participate in activities that require integrating ideas from various sources, including diverse perspectives in their academic work, and discussing ideas with others outside of class" (p. 477). Reflective Learning is a three-item scale that asks "how often students examined the strengths and weaknesses of their own views and learned something that changed their understanding" (p. 477). The Overall Deep Learning Scale is based on a combination of all 12 items. We present the specific items constituting each of the three deep learning subscales and the overall scale in Table 1.

Control Variables

The longitudinal nature of the Wabash Study permitted us to introduce a wide range of statistical controls, not only for student background and precollege traits and experiences, but also for other experiences during the first year of college. Our control variables in the present study were collected during the initial data collection and included the following demographic and precollege variables: a parallel precollege measure of the DIT2 N2-score, sex, race,[4] average parental education in years (converted from an ordinal scale to a pseudo-continuous measure),[5] precollege ACT score or equivalent, precollege political orientation, precollege interest in cognitively-demanding activities (operationalized using the Need for Cognition Scale, Cacioppo et al. 1984), and precollege academic motivation. We provide detailed description of these precollege control covariates in Table 2. We also statistically accounted for the influence that college course-taking patterns has on moral reasoning. These data were collected during the follow-up collection period and included the total number of courses completed in two areas: Humanities, Social Sciences, Education, and Business as well as the Natural Science, Technology, Engineering, and Mathematics (e.g., STEM fields) and Health Sciences. Table 2 contains descriptive statistics for all study variables.

[3] Forty-seven students completed both data collections but failed reliability checks. These students tended to be disproportionately male and students of color. Although the weighting procedure was designed to derive an analytical sample mirroring the population from which it was drawn, it could not address the potential non-response bias that this type of nonrandom missing data might represent.

[4] We recognize the limitation of coding race into a dichotomous measure. By accounting for the correlated or clustered nature of the data (Raudenbush and Bryk 2001), the degrees of freedom available in the models were determined by the number of institutions in the sample ($k = 19$). This strictly limits the number of predictor variables in the regression specification and requires a parsimonious model.

[5] Some have cautioned this common practice in educational research (Long 1997 in Wells et al. 2009) because it may violate the randomly distributed error term assumption.

Table 1 National survey of student engagement deep approaches to learning scales and component items[a]

Scale Items
Higher order learning (alphas = .82, .75)[b]
Analyzed the basic elements of an idea, experience, or theory, such as examining a particular case or situation in depth and considering its components
Synthesized and organized ideas, information, or experiences into new, more complex interpretations and relationships
Made judgments about the value of information, arguments, or methods, such as examining how others gathered and interpreted data and assessing the soundness of their conclusions
Applied theories or concepts to practical problems or in new situations
Integrative learning (alphas = .72, .67)[b]
Worked on a paper or project that required integrating ideas or information from various sources
Included diverse perspectives (different races, religions, gender, political beliefs, etc.) in class discussion or writing assignments
Put together ideas or concepts from different courses when completing assignments or during class discussions
Discussed ideas from your readings or classes with faculty members outside of class discussions
Discussed ideas from your readings or classes with others outside of class (students, family members, co-workers, etc.)
Reflective learning (alphas = .81, .76)[b]
Examined the strengths and weaknesses of your own views on a topic or issue
Tried to better understand someone else's views by imagining how an issues looks from his or her perspective
Learned something that changed the way you understand an issue or concept
Overall deep learning score (alphas = .72, .82)[b]
Includes all 12 items in the scales above

Note: Response options for the *integrative learning* and *reflective learning* scales were: 1 = never, 2 = sometimes, 3 = often, 4 = very often. Response options for the *higher-order learning* scale were: 1 = very little, 2 = some, 3 = quite a bit, 4 = very often

[a] *Source*: Nelson Laird et al. (2008)

[b] First alpha reliability is from Nelson Laird et al. (2008) second alpha reliability is based on the WNSLAE sample

Data Analyses

Our analyses were conducted in two stages. In the first stage we gained a preliminary sense of the relationship between the deep learning scales and the DIT2 N2-scores. These analyses were based on the zero-order correlations between the deep learning scales and both the precollege and end-of-first-year DIT2 N2-scores, as well as the partial correlations between the deep learning scales and end-of-first-year DIT2 N2-score controlling for the precollege DIT2 N2-score. Since these analyses did not bear the major interpretive burden of the study, they did not take into account the clustered or nested nature of the data (as discussed below). However, because such adjustments for nested data typically lead to larger standard errors (Raudenbush and Bryk 2001), we used a more stringent alpha level ($p < .01$) to indicate statistical significance.

The second stage of analysis contained two waves of modeling that were each done, initially including the Overall Deep Learning Scale, and subsequently including the three deep learning subscales (Higher-Order Learning, Integrative Learning and Reflective

Table 2 Descriptive statistics for all variables

Variable	*M*	SD	Minimum	Maximum
Male (reference group = female)	.434	NA	0	1
White (reference group = non White)	.822	NA	0	1
Parental education, measured as the average of the respondents' parental education in years provided the student supplied a response for at least one parent. Response options ranged from did not finish high school (11) to doctorate (20)	15.320	2.259	11	20
Precollege ACT or equivalent, measured as student's ACT score or equivalent SAT/COMPASS score	25.402	4.595	14	36
Precollege political orientation, measured on a five point scale from 1 = far left to 5 = far right	2.870	.867	1	5
Precollege interest in cognitively-demanding activities, measured using the eighteen item Need for Cognition Scale ($\alpha = .89$) (Cacioppo et al. 1984)	3.403	.624	1.222	4.944
Precollege academic motivation, measured using an eight-item scale ($\alpha = .69$). Constituent items included the extent to which students agreed/disagreed that they were willing to work hard to learn the material even if it doesn't lead to a higher grade and the importance of getting good grades	3.542	.554	1.125	5
Humanities, Social Science, Education, and Business coursework	4.513	1.900	0	10
STEM fields and Health Sciences coursework	2.533	1.564	0	6
Overall deep learning	59.597	15.368	5.556	100
Higher-order learning	68.094	19.565	0	100
Integrative learning	53.864	17.430	0	100
Reflective learning	57.813	22.291	0	100
Precollege DIT2-N2 score	33.205	15.060	.284	80.058
End-of-first-year DIT2-N2 score	37.863	15.647	−.958	81.370

Learning). First, we regressed end-of-first-year (spring 2007) DIT2 N2-scores on deep approaches to learning and all control variables (wave 1). Next, we estimated if the effects of deep approaches to learning on DIT2 N2-scores were general (i.e., the same for all students in the sample) or conditional (i.e., differing based on student characteristics). In this step, we first created a series of cross-product variables, consisting of sex (male), race (White), precollege N2-score, and number of courses taken in the STEM fields and health sciences on the one hand and the three standardized deep approaches to learning subscales on the other. We then added blocks of the cross-product terms to the general effects equations specified for wave 1 separately, examining if the cross-product terms significantly increased the amount of variance explained in the criterion variable (Pedhazur 1982). For example, we added the three cross-product terms of male*higher-order learning, male*integrative learning, and male*reflective learning to the wave 1 model specification to see if the effects of the deep approaches to learning subscales differed for men than women. We proceeded in subsequent fashion with the other blocks of cross-product terms.

As noted in the sample description, the Wabash Study first sampled institutions (the primary sampling unit) from which students were sampled. Because students within an institution are more similar than between institutions, it is likely their error terms correlate, violating one of the assumptions of Ordinary Least Squares regression (Ethington 1997; Raudenbush and Bryk 2001). If these correlations (e.g., the intraclass correlation coefficient)[6] are not taken into account, the standard error of the estimates may be incorrect (usually underestimated) resulting in invalid significance tests. We adjusted for the intraclass correlation coefficient by using statistical techniques that adjust for correlated data (Groves et al. 2004). Specifically, we employed the regression option (svy) in the STATA software package that adjusts standard errors of the coefficient estimates. This type of adjustment allowed us to focus on the individual student-level experiences of each predictor with respect to the outcome.

All continuous variables in the model were standardized. This allowed the regression coefficients for all variables to be interpreted as effect sizes. A one unit change in *X* yielded a one standard deviation change in *Y*.

Limitations

Several limitations are worth noting for this study. First, our decision to study students within their nested learning environments curtailed the number of covariates we could include in the analytic models. Working with more limited degrees of freedom resulted in decisions to reduce the number of process variables included in the model, to code race into two discrete categories, and to adopt a similar coding scheme for course-taking behaviors.

Second, we may have been limited by the measure of deep learning approaches as the NSSE overall measure contains only 12 items, while other deep approaches to learning measures contain more items and distinguish between deep and other types of approaches (Biggs et al. 2001; Entwistle and McCune 2004). Future research should consider whether a more robust measure or measures of deep approaches to learning would show stronger relationships even with controls.

Third, our sample consisted of only first-year students. This made it difficult to study certain covariates of interest, namely year in school and depth or quality of co-curricular experience. In addition, we tested these students at only two time points; we cannot speak to stability of change scores over time. Future research should address these limitations by designing multi-year (at least three time points) longitudinal studies of students as they matriculate through college and beyond.

Fourth, we wrestled with including community college students as part of our longitudinal sample. As noted by previous scholars (Cohen and Brawer 2003; Grubb 1996), the institutional mission of community colleges differs from those at other types of colleges; this difference often makes it problematic to include students enrolled at community colleges with other students in multi-institutional studies of college impact. However, research has also shown that many first-year students at community colleges experience cognitive and non-cognitive changes that are quite similar in direction and magnitude to those of their 4-year college counterparts (Pascarella 1999; Pascarella and Terenzini 2005). In addition, we did not want to suggest that a study of first-year students should only

[6] The intraclass correlation coefficient for this analytic sample predicting end-of-first-year N2 score was .188. Failing to adjust for the correlated nature of the data in this sample would artificially lower the standard error of estimates erroneously finding statistical significance.

include the voices of 4-year college students. As a result, we opted to include community college students in our sample and control for student clustering and precollege ACT or equivalent. Future researchers may want to investigate community colleges and their embedded and distinctive educational practices for their effects on moral reasoning development.

Because the focus of this study is on individual student experiences of deep approaches to learning and their relationship to moral reasoning development, we chose not to model institutional-level effects but simply account for the clustered nature of the sampling design. We believe examining institutional-level (i.e., organizational or contextual effects) covariates and their relationship to moral reasoning development would be a fruitful area for further research, as long as there is enough statistical power for examining Level-2 variables.

Results

Table 3 summarizes the results of our preliminary analyses. As the table indicates, the zero-order correlations between the deep learning scales and both the precollege and end-of-first-year DIT2 N2-scores were modest in magnitude. However, there was a clear trend for the associations between deep learning and DIT2 N2-scores to be stronger at the end of the first year of college than the associations between deep learning and precollege N2-scores. Indeed, even controlling for precollege DIT2 N2-scores, the partial correlations between each of the deep approaches to learning scales and end-of-first-year DIT2 N2-scores were all statistically significant, if modest in magnitude. This suggested at least the possibility of a stronger relationship between deep learning experiences and moral reasoning growth during the first year of college.

Table 4 shows the results of the general effects regression estimates. Columns 2 and 3 show the parameter estimates (calculated as effect sizes) and standard errors for the DIT2 N2-score model including the Overall Deep Learning Scale. As Table 4 indicates, the Overall Deep Learning Scale had only a small (effect size = .08) and marginally ($p < .10$) significant influence on end-of-first-year N2-scores when covariates associated with moral reasoning development and deep learning approaches were entered into the model. However, consistent with past literature, women have higher scores on average on the DIT2 than men; students with higher ACT scores (or equivalent) and those who have higher scores on the Need for Cognition pretest demonstrate a higher level of development on the DIT2 in their first year of college (see King and Mayhew 2002, 2004). Consistent

Table 3 Correlations among deep learning scales and DIT2 N2-scores ($N = 1{,}457$)

	Overall deep learning	Higher-order learning	Integrative learning	Reflective learning
Zero-order correlation with precollege N2-score	.104**	.062	.068*	.125**
Zero-order correlation with end-of-first-year N2-score	.193**	.105**	.171**	.187**
Partial correlation with end-of-first-year N2-score, controlling for precollege N2-score	.170**	.087**	.174**	.141**

* $p < .01$; ** $p < .001$

Table 4 Estimated effects of deep learning scales on end-of-first-year DIT2-N2 scores

Predictor	*B*	SE *B*	*B*	SE *B*
Precollege DIT2-N2 score	.550***	.038	.547***	.037
Male	−.166***	.030	−.170***	.027
White	.076	.048	.061	.046
Parental education	.019	.021	.019	.021
Precollege ACT or equivalent	.216***	.033	.224***	.032
Precollege political orientation	.005	.038	.003	.039
Precollege interest in cognitively-demanding activities	.095***	.025	.094**	.028
Precollege academic motivation	.018	.019	.013	.021
Humanities, Social Science, Education, and Business coursework	−.037	.025	−.041	.025
STEM fields and Health Sciences coursework	−.112***	.027	−.108***	.027
Overall deep learning scale	.084†	.043		
Higher-order learning			−.020	.042
Integrative learning			.101**	.033
Reflective learning			.020	.024
R^2	.565***		.568***	

Note: All regression coefficients listed under *B* can be interpreted as effect sizes

† $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$

with the emerging literature on deep approaches to learning, students who have completed a greater number of courses in STEM fields and the health sciences have lower scores on principled moral reasoning in their first year of college (Nelson Laird et al. 2008).

Columns 4 and 5 in Table 4 summarize the model including the three deep learning subscales. Adding the three deep approaches to learning subscales to the base model of control variables, significantly increased the amount of explained variance in the criterion measure from .5598 to .5682 (Wald test $p = .035$). Although Higher-Order and Reflective Learning had no significant unique influence in the full model, we found a small but significant, positive effect for the Integrative Learning subscale ($b = .10$; $p < .01$). We acknowledge this is a small effect and one that some critics may regard as trivial. Given the inclusion of a parallel pretest (which alone explains nearly .50 of the variance in the criterion measure) and that students had been exposed to college-level deep approaches to learning for only seven to 8 months, we argue this small but significant positive effect of Integrative Learning on moral reasoning development should not be dismissed.

Again, the covariates that were significant in the Overall Deep Learning Scale model were statistically significant in the deep learning subscales model. We also examined if shared variance among the deep learning subscales may have contributed to the lack of unique effects on moral reasoning. Using the controls to form a base model, we estimated separate regression models for each of the deep approaches to learning subscales. Analyzing the effects of the deep approaches to learning subscales in separate models did not alter the overall findings. Higher-order Learning had no influence on moral reasoning; Integrative Learning continued to have a statistically significant positive effect on moral reasoning; and Reflective Learning had a positive effect at the $p < .10$ level of statistical significance.

In examining the extent to which the effects of deep approaches to learning on moral reasoning development may differ on the basis of students' characteristics or course-taking

patterns, we found no evidence of such a difference. None of the sets of cross-product terms were associated with a statistically significant increase in the explained variance in end-of-first-year N2-scores. Thus, the estimated unique effects of the deep learning scales on first-year moral reasoning summarized in Table 4 appear to be essentially similar in magnitude for men and women, for White students and students of color, for students with different levels of precollege moral reasoning, and for students irrespective of the number of courses they have taken in the STEM fields and the health sciences.

Discussion

This study takes a small but important step toward understanding the role educational processes play in influencing moral reasoning development among first-year students. Results indicate a modest positive relationship between overall deep approaches to learning and moral reasoning at the end of the first year, even when controlling for precollege moral reasoning. Yet, after controlling for a host of demographic and precollege covariates, certain deep approaches to learning were more conducive to moral growth than others. Further, the relationship between these approaches and moral reasoning development was similar across the first-year students in this sample, at least based on characteristics such as self-identified gender, race, and political orientation.

Results from this study were somewhat surprising in that the overall deep learning measure and two of its constituent scales, higher-order and reflective learning, had weak relationships with moral reasoning development once all controls were introduced, but the subscale capturing integrative learning had a modest, but significant, positive effect. On one hand, this pattern of results is disheartening, given the empirical support suggesting that each deep approach to learning was arguably related to moral reasoning development. On the other hand, the modest correlations and partial correlations between the NSSE deep approaches to learning scales and the moral reasoning scores were promising, even if the relationships diminished when controls were introduced. Further, by controlling for theoretically-based demographic covariates and accounting for students clustering within institutions, we are able to make substantive claims regarding integrative learning and its relationship to moral reasoning development and focus on these claims as evidence for instructing educators interested in helping their students make moral gains in their first-year of college.

That integrative learning shares a significant relationship with moral reasoning development marks this study's most important potential contribution to the literature. This study is the first to examine moral reasoning development as a function of integrative learning, involving students' participating in activities designed to integrate information from varied sources and diverse perspectives. Until now, the relationship between integration and moral reasoning development was, at best, inferred by linking moral reasoning gains to exposure to an integrated curriculum within a distinctive institutional environment (Mentkowski et al. 2000; Mentkowski and Strait 1983; Rogers 2002). By directly measuring integrative learning as an educational process, this study informs educators interested in optimizing learning environments for moral reasoning gains among first-year students: Enact educational practices that teach students how to integrate information from varied sources that draw on diverse perspectives and that motivate students to discuss ideas outside of class.

It is also worth noting that educators should not abandon reflective and higher-order learning approaches based on our findings. They too had modest positive relationships with

end-of-first-year moral reasoning independent of precollege moral reasoning. However, in the final model, their effects were not independent of the effects of the other covariates. This suggests that the effects of higher-order and reflective learning (as well as the overall deep learning scale) overlap with the effects of gender, race, and other background characteristics, as well as students' course-taking patterns. An intriguing follow-up to our study might examine the effects of disrupting the normal patterns of reflective and higher-order learning found among student groups and within particular fields (e.g., pushing more reflective learning in the sciences) and then reexamining the effects of the subscales on moral reasoning.

Turning to covariates of interest, it is not surprising that gender, precollege academic preparation, and precollege interest in cognitively-demanding activities shared significant relationships with moral reasoning, as each has been well-substantiated in the literature (see King and Mayhew 2002, 2004). We were, however, discouraged with the results suggesting first-year students enrolled in more STEM and health courses were less likely to make moral reasoning gains than those with fewer such courses, especially in light of previous research suggesting that results were mixed with regard to disciplinary differences in moral reasoning scores (King and Mayhew 2002, 2004; Snodgrass and Behling 1996; St. Pierre et al. 1990). Perhaps, STEM educators are more concerned with presenting didactic information about chemicals, formulas, and anatomies than with helping students develop frameworks for making meaning of divergent perspectives replete with competing truth claims. We urge instructors of first-year STEM and health science students to present course content and enact practices that challenge students to develop and articulate strategies for grappling with competing hypotheses and points of view. Having wrestled with these issues in their foundational STEM courses, first-year students may be better equipped to adopt similar reasoning strategies when faced with moral dilemmas.

Implications

Several implications emerge from this study. From a research perspective, we were able to account for the nested nature of the data, that is, students in their natural, albeit nonrandom, learning environments. Adopting such strategies for college impact research will become increasingly important for multi-institutional studies, as increasing national concerns for accountability begins to extend its legislative reach to higher education institutions by calling for evidence of student learning.

Also important is how the relationship between integration and moral reasoning development contributes to theory building, informing cognitive-structural approaches to moral reasoning and its ontology (see Kohlberg 1976; Rest et al. 1999). Recent research has suggested that merely exposing students to diverse perspectives may not be enough to disrupt existing schema associated with egocentric frames for understanding justice, as many students may retreat from, rather than work through, the discomfort engendered by confronting unfamiliar perspectives (see Mayhew and Engberg 2010). What may matter more, in terms of inducing moral reasoning development, is teaching students how to successfully work through this discomfort—according to this study's results, this process may involve teaching students strategies for integrating divergent perspectives from varied sources into expanded frameworks for understanding justice as it relates to serving the needs of all people, even those unfamiliar or unknown. For moral reasoning gains to occur among first-year students, educators may need to teach students to integrate information from varied sources and perspectives and then to use the resulting framework as the basis for making moral decisions.

In terms of instruction, it may not be enough to merely expose students to diverse perspectives or varied information sources. What appears more necessary, at least as results from this study suggest, is teaching students how to integrate these perspectives into frameworks that lead to more sophisticated understandings of justice, including its equitable administration. Educators may want to present students with curricular content where arguments are developed and presented from a variety of competing perspectives and subsequently enact practices that encourage students to integrate tenets from these arguments into frameworks for making meaning of the issue at hand. An example of such a practice might include challenging students to develop position papers where competing objectives from multiple perspectives are synthesized into a coherent set of strategies or principles a student then uses as a guide for reasoning about moral dilemmas. Considering a variety of perspectives as a reasoning strategy before ultimately making a decision may help students adopt broader perspectives when faced with moral dilemmas.

At the institution level, assessment findings, including those from the NSSE deep approaches to learning items, are being used to stimulate dialogue about disciplinary differences in teaching and learning practices in hopes of increasing the intellectual climate and faculty emphasis on deeper learning (NSSE 2007, 2009). Disciplinary socialization and tradition, as pointed out by Nelson Laird et al. (2008), is a tremendous challenge to these efforts. However, with established connections between deep approaches to learning and important educational outcomes, like moral development, these efforts have both greater urgency and increased chances of success.

Conclusion

Despite its ubiquitous presence in higher education, very few multi-institutional studies have been designed to explore moral reasoning and the educational processes that lead to its development. This study is one of the first to do so, with its emphasis on examining deep approaches to learning as likely determinants of moral reasoning development. Having found, after controlling for a host of demographic and relevant student characteristics and accounting for the natural clustering of students, that students' participation in activities designed to integrate information from varied sources and diverse perspectives had a unique positive effect on gains in moral reasoning among first-year students, we recommend that educators seeking to improve students' moral reasoning place greater emphasis on such forms of integrative learning.

Acknowledgments The research on which this study was based was supported by a generous grant from the Center of Inquiry in the Liberal Arts at Wabash College to the Center for Research on Undergraduate Education at the University of Iowa.

References

Astin, A. (1993). *What matters in college?* San Francisco: Jossey-Bass.

Barnett, R., Evans, J., & Rest, J. (1995). Faking moral judgment on the Defining Issues Test. *British Journal of Social Psychology, 34*, 267–278.

Biggs, J. B. (1987). *Student approaches to learning and studying*. Hawthorn, Victoria: Australian Council for Educational Research.

Biggs, J. B. (1989). Approaches to the enhancement of tertiary teaching. *Higher Education Research and Development, 8*, 7–25.

Biggs, J. B. (2003). *Teaching for quality learning at university*. Buckingham: Open University Press.
Biggs, J. B., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor Study Process Questionnaire: R-SPQ-2F. *British Journal of Educational Psychology, 71*, 133–149.
Brabeck, M. (1983). Moral judgment: Theory and research on differences between males and females. *Developmental Review, 3*, 274–291.
Burwell, R. M., Butman, R. E., & Van Wicklin, J. (1992). *Values assessment at three consortium colleges: A longitudinal follow-up study*. Houston, NY: Houghton College. ERIC Document Reproduction Service, ED 345 635.
Cacioppo, J. T., Petty, R. E., & Kao, C. E. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*, 306–307.
Cohen, A., & Brawer, F. (2003). *The American community college*. San Francisco: Jossey-Bass.
Crowson, H. M. (2004). Do cognitive motives predict moral judgment development? Unpublished manuscript submitted for publication to the *Journal of Moral Education* on August 9, 2004.
Dewey, J. (1933). *How we think. A restatement of the relation of reflective thinking to the educative process* (revised ed.). Boston: D. C. Heath.
Elmer, N. P., Palmer-Canton, E., & St. James, A. (1997). Politics, moral reasoning, and the DIT: A reply to Barnett et al. (1995). *British Journal of Social Psychology, 37*, 457–476.
Elmer, N. P., Renwick, S., & Malone, B. (1983). The relationship between moral reasoning and political orientation. *Journal of Personality and Social Psychology, 45*, 1073–1080.
Entwistle, N. J. (1981). *Styles of learning and teaching: An integrated outline of educational psychology for students, teachers and lecturers*. Chichester: Wiley.
Entwistle, N. J., & McCune, V. (2004). The conceptual bases of study strategy inventories. *Educational Psychology Review, 16*(4), 325–345.
Ethington, C. A. (1997). A hierarchical linear modeling approach to studying college effects. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. XII, pp. 165–194). New York: Agathon Press.
Fisher, D. G., & Sweeney, J. T. (1998). The relationship between political attitudes and moral judgment: Examining the validity of the DIT. *Journal of Business Ethics, 17*(8), 905–916.
Foster, J. D., & LaForce, B. (1999). A longitudinal study of moral, religious, and identity development in a Christian liberal arts environment. *Journal of Psychology and Theology, 27*(1), 52–68.
Gilligan, C. (1977). In a different voice: Women's conceptions of the self and morality. *Harvard Educational Review, 47*, 481–517.
Gongre, W. M. (1981). *A study of moral judgment, using the Defining Issues Test, for three ethnic groups at Bacone College*. Unpublished Doctoral Dissertation, Oklahoma State University, OK.
Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. Hoboken, NJ: Wiley-Interscience.
Grubb, W. N. (1996). *Working the middle: Strengthening education and training for the mid-skilled labor force*. San Francisco: Jossey-Bass.
Hendel, D. D. (1991). Evidence of convergent and discriminant validity in three measures of college outcomes. *Educational and Psychological Measurement, 51*, 351–358.
Hurtado, S., Mayhew, M. J., & Engberg, M. E. (2003, November). *Diversity in the classroom and students' moral reasoning*. Paper presented at the Association for the Study of Higher Education, Portland, OR.
King, P. M. (2005). Theories of college student development: Sequences and consequences. In M. E. Wilson & L. E. Wolf-Wendel (Eds.), *ASHE reader on college student development theory* (pp. 43–53). Boston: Pearson Custom Publishing.
King, P. M., & Mayhew, M. J. (2002). Moral judgment development in higher education: Insights from the Defining Issues Test. *Journal of Moral Education, 31*(3), 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. XIX, pp. 375–440). The Netherlands: Kluwer Academic Publishers.
Kohlberg, L. (1976). Moral stages and moralization: The cognitive-developmental approach. In T. Likona (Ed.), *Moral development and behavior: Theory, research, and social issues*. New York: Holt, Rinehart & Winston.
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. New York: Cambridge University Press.
Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: SAGE.
Mayhew, M. J., & Engberg, M. E. (2010). Diversity and moral reasoning: How negative diverse peer interactions affect the development of moral reasoning in undergraduate students. *Journal of Higher Education, 81*(4), 459–488.

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

Mayhew, M. J., Seifert T. A., & Pascarella, E. T. (in press). How the first year of college influences moral reasoning development for students in moral consolidation and moral transition. *Journal of College Student Development.*

Mayhew, M. J., Seifert, T. A., & Pascarella, E. T. (2010). A multi-institutional assessment of moral reasoning development among first-year students. *Review of Higher Education, 33*(3), 357–390.

Mayhew, M. J., Wolniak, G. C., & Pascarella, E. T. (2008). How curricular content and educational practices affect the development of life-long learning outcomes in traditionally-aged undergraduate students. *Research in Higher Education, 49*(9), 317–356.

Mentkowski, M., et al. (2000). *Learning that lasts: Integrating learning, development, and performance in college and beyond*. San Francisco, CA: Jossey-Bass.

Mentkowski, M., & Strait, M. J. (1983). *A longitudinal study of student change in cognitive development, learning styles, and generic abilities in an outcome-centered liberal arts curriculum* (Final Report to the National Institute of Education No. 6). Milwaukee, WI: Alverno College, Office of Research and Evaluation.

Murk, D. A., & Addleman, J. A. (1992). Relations among moral reasoning, locus of control, and demographic variables among college students. *Psychological Reports, 70*(2), 467–476.

Narvaez, D., Getz, I., Rest, J. R., & Thoma, S. J. (1999). Individual moral judgment and cultural ideologies. *Developmental Psychology, 35*(2), 478–488.

National Survey of Student Engagement. (2006). *The college student report: 2006 codebook*. Bloomington, IN: Indiana University Center for Postsecondary Research.

National Survey of Student Engagement. (2007). *Experiences that matter: Enhancing student learning and success*. Bloomington, IN: Indiana University Center for Postsecondary Research.

National Survey of Student Engagement. (2009). *Using NSSE to assess and improve undergraduate education: Lessons from the field 2009*. Bloomington, IN: Indiana University Center for Postsecondary Research.

Nelson Laird, T., Shoup, R., & Kuh, G. (2006, May). *Measuring deep approaches to learning using the National Survey of Student Engagement*. Paper presented at the Annual Forum of the Association for Institutional Research, Chicago, IL.

Nelson Laird, T., Shoup, R., Kuh, G., & Schwarz, M. (2008). The effects of discipline on deep approaches to student learning and college outcomes. *Research in Higher Education, 49*, 469–494.

Nevin, K. J., & McNeel, S. P. (1992). Facilitating student moral development through faculty development. *Moral Education Forum, 17*(4), 12–18.

Pascarella, E. T. (1985). College environmental influences on learning and cognitive development: A critical review and synthesis. In J. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 1). New York: Agathon.

Pascarella, E. T. (1999). New studies track community college effects on students. *Community College Journal, 69*, 9–14.

Pascarella, E. T. (2006). How college affects students: Ten directions for future research. *Journal of College Student Development, 47*(5), 508–520.

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

Pedhazur, E. T. (1982). *Multiple regression in behavioral research: Explanation and prediction* (2nd ed.). New York: Holt, Rinehart & Winston.

Quarry, E. B. (1997). *The Defining Issues Test as a measure of moral reasoning of undergraduate students at a Christian liberal arts institution in southern California.* Unpublished Doctoral Dissertation, Biola University.

Ramsden, P. (2003). *Learning to teach in higher education*. London: Routledge Falmer.

Raudenbush, S. W., & Bryk, A. S. (2001). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage Publications.

Rest, J. R. (1979a). The impact of higher education on moral judgment development. Moral Research Projects, Technical Report No. 5. Minneapolis, MN.

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

Rest, J. R. (1987, November 21–24). *Moral judgment: An interesting variable for higher education research*. Paper presented at the Annual Meeting of the Association for the Study of Higher Education, Baltimore, MD.

Rest, J. R. (1988). Why does college promote development in moral judgment? *Journal of Moral Education, 17*(3), 183–194.

Rest, J. R., & Deemer, D. (1986). Life experiences and development pathways. In J. R. Rest (Ed.), *Moral development: Advances in research and theory* (pp. 28–58). New York: Praeger Press.

Rest, J., Narvaez, D., Thoma, S., & Bebeau, M. (1999). DIT2: Devising and testing a revised instrument of moral judgment. *Journal of Educational Psychology, 91*, 644–659.

Rest, J. R., Thoma, S. J., & Edwards, L. (1997). Designing and validating a measure of moral judgment: Stage preference and stage consistency approaches. *Journal of Educational Psychology, 89*(1), 5–28.

Rogers, G. (2002). Rethinking moral growth in college and beyond. *Journal of Moral Education, 31*(3), 325–338.

Sanders, C. E., Lubinski, D., & Benbow, C. P. (1995). Does the Defining Issues Test measure psychological phenomena distinct from verbal ability? An examination of Lykken's query. *Journal of Personality and Social Psychology, 69*, 498–504.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.

Shaver, D. G. (1985). A longitudinal study of moral development at a conservative, religious, liberal arts college. *Journal of College Student Development, 26*(5), 400–404.

Shaver, D. G. (1987). Moral development of students attending a Christian, liberal arts college and a Bible college. *Journal of College Student Development, 28*(3), 211–218.

Snodgrass, J., & Behling, R. (1996). Differences in moral reasoning between college and university business majors and non-business majors. *Business and Professional Ethics Journal, 15*, 79–84.

St. Pierre, K. E., Nelson, E. S., & Gabbin, A. L. (1990). A study of ethical development of accounting majors in relation to other business and nonbusiness disciplines. *Accounting Educator's Journal, 3*, 23–35.

Stepp, J. O. (2002). *The relationship of moral development to critical thinking and family structure of matriculating college students.* Unpublished Doctoral Dissertation, The Southern Baptist Theological Seminary.

Tagg, J. (2003). *The learning paradigm college*. Boston, MA: Anker.

Thoma, S. J. (1993). The relationship between political preference and moral judgment development in late adolescence. *Merrill-Palmer Quarterly, 39*(3), 359–374.

Walker, L. J. (1984). Sex differences in the development of moral reasoning: A critical review. *Child Development, 55*, 677–691.

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

Wells, R., Lynch, C., & Seifert, T. (2009, November). *Analyzing educational expectations: methodological options and their implications.* Paper presented at the Annual Meeting of the Association for the Study of Higher Education, Vancouver, BC.

Whiteley, J. M. (1982). Approaches to the evaluation of character development. In J. M. Whiteley (Ed.), *Character development in college students* (pp. 111–172). New York: Character Research Press.