Moral Judgement Development in Higher Education: insights from the Defining Issues Test

PATRICIA M. KING & MATTHEW J. MAYHEW
University of Michigan, USA

ABSTRACT  This article reviews 172 studies that used the Defining Issues Test to investigate the moral development of undergraduate college students and provides an organisational framework for analysing educational contexts in higher education. These studies addressed collegiate outcomes related to character or civic outcomes, selected aspects of students’ collegiate experiences related to moral judgement development and changes in moral reasoning during the college years as they related to changes in other domains of development. Findings suggest that dramatic gains in moral judgement are associated with collegiate participation, even after controlling for age and entering level of moral judgement. Although many studies used gross indicators of collegiate context (e.g. institutional type or academic discipline), studies that examine specific collegiate characteristics and educational experiences are better suited to identifying factors that contribute directly or indirectly to changes in moral judgement during the college years. Implications for student development practice and future research are discussed.

The Defining Issues Test (DIT) has been widely used with samples of undergraduate college students to investigate a broad range of moral issues. Consequently, this instrument has played a major role in shaping our understanding of the development of moral judgement among college students. In their major review (over 2600 studies) of the effects of college on students, Pascarella and Terenzini (1991) devoted a whole chapter to moral development, and noted that there exists impressive evidence of moral development in the college years, both in terms of the sheer number of studies conducted and in the diversity of samples tested. As they are currently updating this book (Pascarella & Terenzini, in preparation) we wished to complement rather than replicate their efforts, so our guiding question for this essay was, “How has the DIT been used in higher education?” Starting here allows for a broader vantage point from which to examine how the DIT has informed our understanding of moral development among college students. As Rest (1984) and others in this JME Special Issue point out, (Bebeau, 2002; Rogers, 2002; Thoma, 2002 issue), the domain of morality is much broader than the moral judgement component that the DIT is designed to measure; this review is limited to one aspect of moral development, moral judgement.
This student population is important for several reasons. First, it is common for both traditional- and non-traditional-age students to enroll in college at times in their lives when they are making important life transitions, many of which have moral implications. This is often accompanied by a new readiness to examine the moral dimension of their lives in preparation for their new life roles. Secondly, most American colleges and universities embrace a mission that arguably includes moral development (such as preparation for citizenship, character development, moral leadership, service to society). Most colleges offer courses in religion, ethics, or what Gaurasci (2001) calls “the democratic arts”, and most offer opportunities for students to participate in community service and service learning projects, all of which have a moral dimension. Thirdly, college graduates often take leadership positions in both their employment settings and in their communities, positions in which they make decisions affecting the lives of others. Thus, there are many reasons to examine moral development in the college years, whether from an individual perspective (e.g. documenting the evolution of moral development among adults attending college), from a college cultural perspective (e.g. examining moral experiences across different institutional subcultures) or from a societal perspective (e.g. assessing the effectiveness of colleges in preparing citizens and workers for their societal roles). Clearly, it is important and valuable to understand how a major assessment tool has influenced our understanding of these important questions.

The impact of the DIT is reflected in the sheer number of studies in which it has been used with college samples. In preparation for writing this essay, we reviewed over 500 published articles, conference presentations and dissertations in which the DIT was used with college students. From this search, we discovered that many of the articles that used both the DIT and a college student sample were not designed intentionally to investigate the moral judgement of college students. Instead, many studies appeared to use college students as convenience samples or as proxies for reasonably bright young adults, but with no particular emphasis on college as an educational context nor on college students as a purposeful sample. As a result, we chose to include in this review only those studies that were intentionally designed to investigate the moral development of undergraduate college students. We also chose to delimit our review to US college samples because a discussion of international students and contexts was beyond the scope of this article. In addition, we omitted samples of college students that were used only to contrast college students with professional samples, as the use of the DIT in the professions is discussed by Bebeau (2002) in this issue. Last, with the exception of original pieces written by James Rest (Rest et al., 1974; Rest, 1979a, 1979b), we chose to limit our review to articles published since 1980, as many of the findings from Rest’s early articles (1974–1979) are subsumed and/or replicated by more recent studies. Accordingly, the 172 studies reviewed here used the DIT to measure the degree to which students achieved intended collegiate outcomes related to character or civic outcomes, studies that focused on selected aspects of students’ collegiate experiences related to moral judgement development (i.e. curricular and co-curricular activities), and studies that documented changes in moral reasoning during the college years as they related to changes in other domains of development.
We began our investigation of how the DIT had been used in collegiate studies of moral development by identifying the types of topics that have been addressed using the DIT and the kinds of variables that have been addressed in these studies. For example, differences in DIT scores by year in college, academic discipline and institutional type have been frequently researched. For each major topic included here, we ask, “What general conclusions can we draw from this body of research?” It is beyond the scope of this paper to provide a detailed review of the findings for each study and each major variable; instead, we offer a general review, and focus on those variables that are distinctive to higher education, or that are of particular interest to educators who work with college students or researchers investigating moral issues.

**Moral Judgement Development in Higher Education**

The purpose of this section is to examine how the DIT has been used to establish the relationship between the development of moral reasoning and participation in higher education. Intentionally or unintentionally, moral development is an outcome of higher education, at least as measured by the DIT (Rest et al., 1974; Rest, 1979a,b, 1987, 1988; Gongre, 1981; Whiteley, 1982; Mentkowski & Strait, 1983; Hood, 1984; Kitchener et al., 1984; King et al., 1985; Rest & Thoma, 1985; Shaver, 1985; Gfeller, 1986; Shaver, 1987; Bouhmama, 1988; Buier et al., 1989; Iceman et al., 1991; Paradise & Dejoie, 1991; Burwell et al., 1992; Jeffrey, 1993; Thoma & Ladewig, 1993; King & Kitchener, 1994; Quarry, 1997; Foster & LaForce, 1999; Loviscky, 2000; Mentkowski et al., 2000; Stroud, 2000; Cummings et al., 2001). With that said, there are only two studies included in this review that either found no relationship between formal education and moral reasoning development (Galotti, 1988) or failed to report differences in moral reasoning by formal education level (Mustapha & Seybert, 1989).

**Longitudinal Studies**

A series of longitudinal studies have been used to investigate the relationship between moral development and formal education by examining the effects of age and education on the development of moral reasoning. These studies suggest that collegiate experiences do promote moral development; more specifically, during college students tend to decrease their preference for conventional level reasoning and increase their preference for postconventional moral reasoning (Rest, 1979b; Whiteley, 1982; Mentkowski & Strait, 1983; King et al., 1985; Shaver, 1985; Shaver, 1987; Burwell et al., 1992; King & Kitchener, 1994). Rest et al. (1999) argued for the use of the label “postconventional” to replace “principled” moral reasoning; we have done so here as well.) Several longitudinal studies have attempted to track the development of moral reasoning by testing students at two times during their collegiate experience, at the beginning of the freshman year and at the end of their senior year (Shaver, 1985, 1987; Burwell et al., 1992; Foster & LaForce, 1999). Other longitudinal studies
have charted this development by testing students at multiple times during their college years (Rest, 1979b; Whiteley, 1982; Mentkowski & Strait, 1983; King & Kitchener, 1994). Rest (1979b) and Rest and Thoma (1985) also used longitudinal data to examine the relationship of moral judgement development to formal education. They tracked the course of moral judgement development of participants from the end of high school to 6 years beyond high school; some attended college and others did not. At the third time of testing, Rest (1979a) found that the course of development for the 38 college students was different from the 18 participants not in college: for those attending college, DIT scores continued to increase; for those not in college, scores were stable. At the fourth time of testing, Rest and Thoma (1985) regrouped the population into low-education (less than two years of college) and high-education (more than two years of college) categories. While both groups showed increases in P scores 2 years after leaving high school, four years later the 23 students with two or more years in college were still showing gains while the 13 less-educated subjects were not (Rest & Thoma, 1985). Because age and education are often confounded in student populations, especially among traditional-age college students, it is important to control for age to test for the impact of college. The design of this study (use of a same-age non-college comparison group and a longitudinal design) provides a model research design for researchers interested in untangling the influence of education from that of maturation on moral development. From this study we can conclude that participation in higher education makes a substantial contribution to development in moral judgement beyond that attributable to age alone.

Other studies have also documented the effect on moral judgement of participation in higher education after controlling for age (Rest, 1979a, 1987, 1988; Kitchener et al., 1984; King et al., 1985; Rest & Thoma, 1985; Gfellner, 1986; Paradice & Dejoie, 1991; Cummings et al., 2001). From these studies, we can conclude that formal education makes a unique contribution to moral reasoning in that during college, students are more likely than non-students to use postconventional moral reasoning.

Pascarella and Terenzini (in preparation) came to a similar conclusion, that the observed increase in postconventional moral reasoning in college “appears to be substantially greater in magnitude than that due merely to maturation and cannot be attributed solely to initial differences in moral reasoning, intelligence, or social status between those who attend and those who do not attend college” (p. 5). This collection of studies provides strong evidence that student participation in higher education is associated with gains in moral development during the college years.

Moral Judgement Development by Ethnicity

There has been considerable interest among scholars and educators about the applicability of theories of college student development that were normed on white samples to students of other racial/ethnic backgrounds (e.g. McEwen et al., 1990; Evans et al., 1998), yet few studies have used the DIT to explore whether moral judgement differs by race or ethnicity among US college students (Gongre, 1981;
Moral Judgement Development in Higher Education

We identified only two studies for which the primary purpose was to investigate differences in moral judgement by race or ethnicity (Gongre, 1981; Locke & Tucker, 1988). The others reported differences in moral judgement by race or ethnicity as they related to other research questions. Gongre (1981) examined the differences between moral judgement of 15 black American, 46 Native American and 53 white students at Bacone College, and found no differences between the three groups on the DIT. This study may be unique in the composition of its sample, as it is the only one in this review that explicitly sampled Native Americans. Although differences were reported among the tribes sampled, information is not provided regarding the sample sizes for the tribes, subgroup means, nor the direction of the difference.

Locke and Tucker (1988) tested black and white students to determine their reactions to moral dilemmas when the race of the central character in the dilemma was changed. They assigned 232 graduate and undergraduate students to one of two groups: the first group received the standard form of the DIT and the second group received a “revised DIT” in which the protagonists in the stories were identified as black. The postconventional moral reasoning scores between the 109 white students taking the standard version of the DIT and those of the 103 white students taking the revised version were not significantly different. However, the 12 black students taking the revised version of the DIT scored significantly lower than the eight black students taking the standard version. This finding is provocative, and merits further exploration. Unfortunately, the small size of the black sample precludes making a strong conclusion about the research question asked, or whether, as the authors suggested, the racial manipulation triggered an emotional reaction among the black students that affected their DIT scores.

The remaining studies that address differences in moral judgement by race or ethnicity do so as supplementary rather than primary questions under investigation (Murk & Addleman, 1992; Johnson et al., 1993; Loviscky, 2000). For example, in his examination of the construct validity of moral judgement as an alternative predictor of job performance, Loviscky (2000) tested 238 undergraduate and graduate students. In a post hoc analysis of DIT scores by ethnicity, he found that students who indicated their ethnicity was other or white had significantly higher DIT scores than black, Hispanic or Asian students, and that the P-scores for Black students were significantly higher than those for Hispanic or Asian students. Unfortunately, sample sizes were not reported by subgroup for this study; however, based on sample percentages reported, we calculated his sample sizes as follows: white ($n = 192$); Asian ($n = 22$); black ($n = 11$); Hispanic ($n = 7$); other ($n = 6$).

Research investigating the relationship between race and ethnicity and moral judgement development measured by the DIT is underdeveloped, and no clear pattern of results is yet available. That few studies have examined the relationship between race and ethnicity and the development of moral judgement is surprising given the racial and ethnic composition of US colleges and universities and the strong interest in diversity-related issues in higher education. Perhaps researchers and educators have a priori knowledge that the DIT was normed using white
samples as norms, and for this reason have avoided its use as a metric of differentiation between ethnic groups when investigating moral judgement development. Or perhaps educators fear that the potential for misusing the DIT could lead to results that may further marginalise under-represented ethnic groups within higher education. Whatever the reason, the dearth of studies makes it impossible to determine whether there are differences by ethnicity on the DIT.

**Collegiate Contexts that are Conducive to Moral Development**

What types of educational experiences stimulate moral development among college students? What are the characteristics of collegiate learning environments that promote development in moral judgement? We turn next to these questions.

As shown above, significant growth in the use of postconventional moral reasoning does occur in college, and this growth is not attributable to general maturation. There are many possible reasons why this might occur: the general intellectual milieu of colleges and universities that fosters the exchange of ideas, exposure to multiple perspectives regarding social issues, academic values of truth-seeking and careful reasoning, or institutional values of academic integrity and personal responsibility. It is also reasonable to assume that some contexts are more effective than others in promoting moral development. For example, those in which the moral dimensions of community life are explicit, where students are encouraged to wrestle with moral dilemmas, and where they are exposed to and encouraged to practise postconventional moral reasoning would be expected to be more effective in promoting postconventional moral reasoning than those without such opportunities and those in contexts that are not known for prosocial behaviours and attitudes. In the next section, we review studies that reflect four types of collegiate contexts and their relationship to moral reasoning. We start at a general level by looking at moral judgement development across institutional types, then consider how the development of moral judgement might differ by academic discipline. Next, we examine interventions that are explicitly designed to promote moral development, and conclude this section with a summary of particular collegiate experiences and their relationship to moral judgement.

**Distinctive Institutional Contexts**

Higher education in the United States is distinguished by the wide variety of types of collegiate institutions, which vary widely across a range of factors, including size, public or private control and mission. Major institutional types range from large public universities with multiple constituencies and missions to small private colleges serving distinctive student subgroups with distinctive missions. Accordingly, institutional type is a common research variable in higher education research, including research on moral judgement (Shaver, 1985, 1987; McNeel, 1991, 1994; Burwell et al., 1992; Good & Cartwright, 1998; Pascarella & Terenzini, 1991, in
preparation). Based on these studies, it appears that the environment of liberal arts colleges tends to be more conducive to fostering the development of moral reasoning than that of other types of colleges and universities (Pascarella & Terenzini, 1991, McNeel, 1994; Good & Cartwright, 1998, in preparation). For example, Good and Cartwright (1998) compared levels of moral judgement development among undergraduate university students attending a state university, a Christian liberal arts university and a Bible university. They found no significant differences in postconventional moral reasoning among the freshman students across institutions; however, senior students at state universities and at Christian liberal arts colleges showed higher levels of postconventional moral reasoning than senior students at Bible colleges.

Shaver (1987) also compared the DIT scores of students attending a Christian liberal arts college and a Bible college, and did so using a longitudinal study. He found that at the time of entry and at the time of exit, postconventional moral reasoning scores for Christian liberal arts students were significantly higher than those for Bible college students. Similarly, in their secondary analysis of data reported by Rest (1979a) on the DIT P-scores of college students, Pascarella and Terenzini (1991) found that moral reasoning differed significantly by institutional type, with students from church-affiliated liberal arts colleges scoring the highest, followed by those at public research universities, two-year colleges, private liberal arts colleges, private universities and public comprehensive universities, respectively. McNeel (1994) also conducted a meta-analysis of college effects on postconventional moral reasoning by college type. Using 22 samples made up of students from seven liberal arts colleges, three universities and two Bible colleges; he reported large effect sizes for liberal arts colleges, large or moderate average effect sizes for universities, and no effect or a moderate effect size for the two Bible colleges. In other words, the only consistently large effect size was found among samples of students attending liberal arts colleges. This is consistent with other studies reporting that liberal arts colleges are more effective than their counterparts in promoting development in moral judgement.

These studies suggest that the development of moral reasoning is affected by the collegiate context. This is a very gross indicator, as “context” is a very general factor that might include a wide variety of more specific factors that are more directly related to change in moral reasoning. Further, there is great variability among institutions that share a similar “type”. For example, some public universities work from an explicit value framework and encourage and expect students to discuss their values and ground their decisions in a value-based framework; others discourage initiatives that suggest religious values, holding separation of church and state as the higher value. Students attending these similar-type institutions would probably have very different collegiate experiences. Nevertheless, findings of institutional differences do provide the impetus to look more closely at more specific features of the collegiate environment. Some of these are discussed below in the section on enriching experiences.
Disciplinary Contexts

The DIT has been used to measure differences in the moral reasoning of college students across academic disciplines (Zeidler & Schafer, 1984; St Pierre et al., 1990; Icerman et al., 1991; Paradice & Dejoie, 1991; Jeffrey, 1993; Ponemon & Gabhart, 1994; Snodgrass & Behling, 1996; Cummings et al., 2001). Variability of moral reasoning scores within certain disciplines has also been observed (Icerman et al., 1991; Paradice & Dejoie, 1991; Jeffrey, 1993).

Several studies have attempted to measure differences in moral reasoning between academic disciplines (St Pierre et al., 1990; Snodgrass & Behling, 1996), yielding inconclusive results. For example, St Pierre et al. (1990) found that accounting majors and students majoring in other business disciplines (i.e. finance, information systems, hotel/restaurant management, management, marketing and international business) showed lower levels of postconventional moral reasoning than did students in psychology, maths and social work. Snodgrass and Behling (1996), by contrast, found no significant differences in the moral reasoning levels between business and non-business majors (i.e. arts and humanities, social sciences, natural sciences and undeclared).

Other studies have examined the development of moral reasoning of students majoring in disciplines that vary in their social orientation (St Pierre et al., 1990; Paradice & Dejoie, 1991). For example, Paradice and Dejoie (1991) found that management information system (MIS) majors generally used higher levels of postconventional reasoning than did non-MIS business majors. They found higher DIT scores among MIS majors, and concluded that these students were more socially orientated than students enrolled in other business majors.

Societal concern with unethical business practices (e.g. insider trading, deceptive advertising, tax fraud, tax evasion) led researchers to investigate the development of postconventional moral reasoning across a series of studies within the business community (St Pierre et al., 1990; Icerman et al., 1991; Paradice & Dejoie, 1991; Jeffrey, 1993; Ponemon & Gabhart, 1994; Snodgrass & Behling, 1996). For example, Jeffrey (1993) examined differences in level of moral development among lower division business students, and found that accounting students had higher postconventional moral reasoning scores than non-accounting business students. Moreover, findings from this study show that senior students in each business major showed higher levels of moral reasoning development than entering students in each major, and the size of the differences was constant across business majors—indicating that the finding of differences in moral reasoning between business majors is not a result of self-selection into a given major. Similarly, Paradice and Dejoie (1991) found that MIS majors (juniors and seniors) used higher levels of postconventional reasoning than did non-MIS business majors. They do not, however, identify specific factors that might contribute to these differences, one of which may be self-selection into specific majors.

Several researchers have proposed disciplinary-specific adaptations of the DIT by grounding the content of the dilemmas in discipline-specific contexts or issues. Specifically, Westbrook (1994) developed a DIT-like test for journalism students
and professionals (called the “Journalist’s Instrument”); Loviscky (2000) developed a “Managerial DIT” (MDIT) to measure the development of moral reasoning for management students and professionals; Lampe (Lampe & Walsh, 1992; Lampe, 1994) developed the “Survey of Educator Ethics Opinions” to measure the moral development for preservice and practicing teachers; and Zeidler and Schafer (1984) used the “Environmental Issues Test”, developed by Iozzi (1978), to measure moral reasoning related to environmental issues. A detailed review of these instruments is beyond the scope of this article. However, they are mentioned here because they highlight the importance of issues of context when making educational decisions about students’ moral reasoning. Some educators find it unwise to extrapolate from dilemmas of a general moral nature to those that are in the context of their field of study, and are less persuaded by psychometric data on national samples of students than by data from specific disciplinary contexts. The researchers who developed these new instruments have attempted to address this issue by retaining the structural format of the DIT, but by using discipline-specific moral dilemmas and items. Whether these adapted instruments will prove to be valid measures of moral judgement will depend on the results of future research.

To date, research on the development of moral judgement within academic disciplines has yielded results that are inconclusive and therefore provide little insight into the nature of the specific characteristics within disciplines that are the most conducive to growth in moral judgement. Although it is understandable that educators and researchers alike are curious about differences by academic discipline, it may be that this by itself serves as only a very preliminary unit of analysis. With that said, understanding the specific content and curricular approaches that make up any given academic discipline is useful for conducting interventions that promote growth in moral reasoning. We turn next to specific kinds of certain collegiate experiences (e.g. ethics courses, community service projects, pedagogies that invite the consideration of moral controversies) that may be more directly related to moral judgement development, and that thus provide more fruitful avenues for research on pedagogies that foster moral judgement.

**Intervention Studies**

The DIT has been used in a number of studies that investigated the effects of participation in educational experiences that were designed intentionally to promote the development of moral reasoning among college students. These vary widely in content and focus, including general education courses (Mustapha & Seybert, 1989, 1990), ethics courses or courses with an ethics component (Armstrong, 1993; Ponemon, 1993; Boss, 1994), social diversity courses (Adams & Zhou-McGovern, 1990, 1994), a freshman colloquium on psychosocial issues (Tennant, 1991), participation in service learning or community service programmes (Boss, 1994; Gorman et al., 1994) and an outdoor education programme (Smith & Bunting, 1999). A much more ambitious approach was the Sierra Project (Whiteley, 1982), an intervention designed to promote character development in students, combining
residential, academic and personal elements, and testing students several times a year over several cohorts.

Given this remarkable variety of approaches, it is especially noteworthy that virtually all these approaches were effective in promoting moral judgement; exceptions were reported by Ponemon (1993) and Tennant (1991). Another distinctive feature of this group of interventions is that most took place in the context of one-term courses, including three that included direct instruction in ethics. However, the results are mixed for effectiveness of the ethics courses: significantly higher moral judgement scores were reported for the intervention samples by Armstrong (1993) using a sample of accounting students, and by Boss (1994; described below), but not by Ponemon (1993), who also studied accounting students.

Boss’s (1994) study is noteworthy in that it controlled for class size, instructor, class exercises and text used across two sections of an ethics class. Boss found that the ethics curriculum and discussion of moral dilemmas and moral development was effective only for the ethics class, whose members also were also required to complete 20 hours of community service work “that involved working directly with people in need” (p. 187) and to keep a journal of their experiences. DIT scores for these students increased, while those of the control group remained stable. Further, 51% of the experimental group and only 13% of the control group used primarily postconventional moral reasoning at the post-test. A similar design (but with fewer controls) was used by Gorman et al. (1994), who tested students enrolled in two courses, “Perspectives on Western Culture”, and “Person and Social Responsibility”. The latter included field projects that “put the students into direct contact with examples of social injustice” (p. 426) and opportunities to reflect upon and discuss their field experiences. Post-test DIT scores were higher for students in the section with the service component.

Adams and Zhou-McGovern (1994) found that courses on social diversity and social justice could also be effective in promoting moral judgement. The course they studied focused on racism, anti-Semitism, sexism, homophobia and disability oppression. By the end of the course, DIT scores had increased significantly.

It appears that courses on topics with a less obvious link to moral education can also provide effective venues for promoting moral judgement. Other effective courses emphasised decision-making and active learning; these components were central to the structure of both the integrated general education curricula tested by Mustapha and Seybert (1989, 1990), the outdoor education course studied by Smith and Bunting (1999) and Tennant’s (1991) psychosocial course for entering freshmen.

Abdolmohammadi et al.’s (1997) study was not an intervention per se, but employed a strategy that those designing interventions might find informative. They administered the DIT to 301 students near the end of a business ethics course; they then randomly assigned these students to three-member groups and asked them to come to a consensus response for each of the questions on the DIT. In the individual administration, P-scores of female students were significantly higher on the DIT ($M = 49$) than were the scores of the male students ($M = 38$). However, in the group administration, the scores of the women declined ($M = 44$) while the
scores of the men increased significantly \((M = 43)\). These results reveal differences in individual versus group decision-making and gender-related effects of group interactions on reasoning.

Two major institution-wide studies that include measures of character development have also employed the DIT and inform our understanding of the development of moral judgement in the college years. The first of these was the Sierra Project study conducted at the University of California-Irvine in the early 1980s, providing one of the most comprehensive longitudinal studies of character development available at that time. Successive cohorts of freshmen were tested on an array of instruments, including the DIT. Their DIT scores increased significantly over time, and were not attributable to growth in other dimensions of development (Whiteley, 1982).

Secondly, the faculty and staff at Alverno College have been engaged in a massive research effort to evaluate adaptable elements of their ability-based curriculum for deep and durable learning in college and beyond; a major volume summarising this effort has been published recently (Mentkowski et al., 2000). They found that students’ growth in moral judgement on the DIT (Rogers, 2002 this issue) was related to progress through Alverno’s developmentally sequenced curriculum, which includes but is not limited to participation in the Valuing in Decision-Making ability area taught within the disciplines. Analysis of student interviews showed which elements of the curriculum and of the campus culture helped broaden their moral perspective. Further, the authors found that students “came to appreciate and understand differing values because they were repeatedly asked to examine and discuss them” (p. 130).

In summary, the majority of studies in this section did report significant increases in moral judgement after the intervention. However, several employed research designs that do not allow the reader to confidently attribute the increases to the intervention _per se_. Specifically, intervention studies designed to promote moral development would be strengthened by meeting experimental research design criteria: carefully selecting control or comparison groups, following strategies that allow for the control of selection effects, providing detailed descriptions of the major features of the intervention to readers, designing conceptually grounded interventions, and testing for stability in change scores after the post-tests.

**Effects of Other Collegiate Contexts on Moral Development**

Several studies have examined aspects of specific collegiate experiences and their relationship to DIT score. Rest (1979a,b, 1986, 1987, 1988) initiated this line of study with data from his (Rest, 1979b) longitudinal sample of 59 students who were re-tested twice in two-year intervals. Of these 38 had attended college, 18 had not and three reported ambiguous information. DIT scores increased for both college and non-college students for the first two years out of high school, but at the third testing the DIT scores of the students still in college were still increasing, while those for the non-college cohort had decreased. Further, when asked to identify experiences they thought had influenced their moral thinking, a wide range of
experiences was cited. One of the few that was cited by those who showed a greater rate of increase than those not citing it was “spending more time contemplating issues”, suggesting that reflection rather than instruction or other specific experiences is a key factor in promoting moral thinking. These preliminary results were examined and explored in more depth in Rest and Deemer (1986).

More recently, Thoma and Ladewig (1993) conducted a study that examined a particular aspect of students’ experiences: they hypothesised that the quality of students’ peer relationships would affect their moral judgement development during college. Using a sample of 156 college students, they found that both DIT scores and the number of close college friends increased for each class level from freshman to senior years. Students with close friendships who had multiple independent friendship groups (that is, those whose friends were not necessarily friends with one another) had higher DIT scores than other students. These findings suggest that diverse friendships may provide a context for challenging students to consider issues of fairness from different perspectives. Alternatively, it may be that those who seek multiple friendships are more open to developmental challenges. Although the links between types of friendships and development of moral judgement have not yet been explicated, this line of research bears much promise. (For a review of the literature on friendship and moral judgement, see Derryberry & Thoma, 2000.)

Two studies have examined the influence of membership in Greek organisations on moral development. Sanders (1990) administered the DIT to male freshmen at the beginning of the autumn term (n = 195) and nine weeks later at end of the term (n = 101). Independents scored significantly higher than fraternity members on both the pre- and post-test; P-scores for both groups are remarkably low, M = 10.0 and 8.6, respectively [1]. Cohen (1982) tested fraternity and sorority members in various leadership positions (23 executive committee members and 111 presidents) and 141 other individual members. She found no differences by membership category, gender or year in school; all had mean P-scores in the high 30s. Cohen interpreted these findings by suggesting that these students “do not seem to be having the upending experiences necessary to make the transition from conventional to post-conventional thinking” (p. 328). These two studies substantiate Derryberry and Thoma’s (2000) observation that “although the college data are supportive of Kohlberg’s theory and suggest an influence of the college environment on moral thinking, it is surprising how little we know about exactly what contributes to this shift” (p. 14). Clearly, this is an area that merits more attention.

Last, an experiment by Reall et al. (1998) compared the moral reasoning of 118 upper-division business students in a non-competitive situation (using the DIT) and in a competitive game (using a Reasoning List based on the DIT that yields the ‘moral reasoning during competition’ [MRC] score). The average P-score was low for college students, 37 (see [1]). The level of moral reasoning exhibited during competition was significantly lower during the competitive game in this study, with the modal level of reasoning being at the preconventional level. The authors concluded that the acceptance of business norms that reflect lower stage reasoning, e.g. ‘that’s the way it works so I’ll do it that way too’ (p. 1209) contributed to the lower scores in a competitive context. However, the comparability between the
MRC scores and the DIT has not been established sufficiently to say whether this discrepancy is the result of the context or the measurement tool. Nevertheless, this study raises provocative questions regarding the effects of collegiate contexts that are competitive, especially among subgroups with P-scores that are below average among college samples.

These studies capture a few of the wide variety of contexts in which college students have experiences that may affect their moral development. Some collegiate environments may foster close friendships with diverse others and therefore foster moral perspective-taking, while others may foster close friendships that reinforce stereotyping or perpetuate non-moral values and behaviours; each in turn can lead to higher or lower moral judgement scores. The next section looks specifically at the links between moral judgement and moral behaviour.

The Relationship between Moral Judgement and Moral Behaviour in College Students

Many collegiate institutions purport to teach their students to be good citizens, provide moral leadership and teach democratic values and decision-making skills that students can use in their professions and in their communities after graduation. As a result, many who study institutional impact and effectiveness are keenly interested in the link between moral judgement and moral behaviour. The link between moral judgement and moral behaviour is critical because, as Thoma (1994) pointed out, understanding moral action may be seen as the “acid test” (p. 199) of the usefulness of research on morality. Fortunately, several studies have been done that examine this relationship. Comprehensive reviews of this literature that include detailed conceptual analysis are available elsewhere (see especially Blasi, 1980; Thoma & Rest, 1986; Rest et al., 1986; Thoma, 1994). The current review is limited to DIT studies conducted with college student samples. It also excludes those studies examining factors that might affect moral judgement (e.g. comprehension, situational characteristics), but for which the observed behaviour (and dependent variable) was only a change in DIT scores. These studies are included in the final section.

We identified 10 studies that fit these criteria. Of these, two used creative experimental designs. Noting that “it is often easier to tell the truth when it is obvious what the truth is” (p. 41), Brabecck (1984) examined the act of whistle-blowing under conditions that made it more difficult to ascertain the truth. She administered the DIT to a sample of 32 juniors and seniors in an introductory counselling class; this yielded a postconventional group (those whose P-scores were above 50) and a conventional group (with scores 50 and below). Half of each group was then assigned randomly to one of two treatments. All were asked to help pilot test questions ostensibly written for a text by the investigator; the chapter contained two factual errors. In Treatment I (n = 13), the errors were pointed out by a peer-confederate from an article in the American Psychologist; in Treatment II (n = 12), the source of the contradictory evidence was ambiguous. No confederate was present for the control group (n = 7). Participants who brought either error to
the investigator’s attention were scored as whistle-blowing. Of those in the conventional group only 8% \((n = 1)\) blew the whistle, while 54% \((n = 7)\) of those in the postconventional group did so; none in the control group blew the whistle. Although this sample was small, the trend was clear: students in the postconventional group were significantly more likely to call a potential error to the attention of the investigator than were those in the conventional group.

The second innovative design was used by Ponemon (1993). The purpose of this study was to evaluate the effectiveness of a one-term case-orientated ethics intervention within an auditing course taken by 126 seniors and graduate students in accounting using a pretest–post-test, experimental–comparison group design; the comparison group received several weeks of instruction on the AICPA Code of Professional Conduct. No significant increases in DIT scores were found for either group of students or either treatment; P-scores remained low for college samples \((M = 38; SD = 12; \text{see [1]})\). An additional measure of moral behaviour was used, as follows. Students were asked on the syllabus and at the beginning of the class to help pay for copies of the course lecture notes (114 pages; $11.40) due to financial constraints on the department copying budget. Payment was optional, and this served as the dependent behavioural variable; not paying was known as “free-riding”. Median payments were between $4.50 (ethics class) and $5.00 (comparison class). Dividing the obtained DIT scores into quartiles, significant differences in payment amount were observed: average payments received from students in the first, second and third quartiles were $2.08, $5.92 and $8.52, respectively. However, the average payment by those in the fourth quartile (i.e. those with the highest DIT scores) was $2.95. Thus, fee payment was found to be related systematically to DIT score, but not in a linear fashion. The pattern of increasing payments from the first to third quartile is consistent with prior research on the judgement/behaviour relationship. However, both the overall low level of paying the fee and the break in the pattern by those with the highest scores raises further questions about the nature of this relationship.

In a debriefing survey, 43% reported that free-riding was unethical. Only 19% indicated they would free-ride, but actually over 67% did not pay the fee at all or in full. In a series of items related to the fee-paying decision, items corresponding to early stages of moral reasoning (preconventional or early conventional) were prevalent. These findings show that many factors affect moral behaviour, and that even reporting an act to be unethical is not a good predictor of acting in morally consistent ways. Whether this is related to the fact that this sample had relatively low P-scores remains to be seen.

Another study of academic responsibilities was conducted by Cummings et al. (2001); they examined self-reported propensity to engage in academic misconduct among teacher education students enrolled in an educational technology course. The average P-scores were also low (36.6) and they also used a quartile split (a high P-score was defined as 47 or above) to analyse the data. Students with higher P-scores reported that they engaged less frequently in academically dishonest behaviours. Whether they actually had done so is not known, a question that is made more salient in light of the discrepancies found in the Ponemon (1993) study.
between self-reported and actual behaviours. This suggests that the use of indirect measures such as used in this study are less reliable as indicators of moral behaviour.

Malinowski and Smith (1985) conducted an experiment with 53 male college students to examine the relation to cheating of moral judgement and other variables thought to affect cheating behaviour. The DIT was administered in a class, and students’ participation in an seemingly unrelated laboratory study of attention–concentration was used to assess cheating behaviour. They found that although most (77%) cheated on at least one trial, the higher the DIT scores, the less cheating occurred. The higher the P-score, the fewer the number of trials on which students cheated, the fewer the number of seconds by which they inflated their scores, and the later in the process cheating began (if at all). DIT scores were also used to group the students into Stage 4 and Stage 3 groups. Stage 3 subjects cheated more and began cheating sooner. Of the 12 men who did not cheat, 11 were in the Stage 4 group. This study was conceptually grounded in Kohlberg’s theory and in Rest’s emerging research on the DIT, was carefully designed to take a variety of potentially confounding variables into account, and helps explicate the role of moral judgement as well as situational variables in determining moral behaviour.

Four studies (Taylor et al., 1985; Cartwright & Simpson, 1990; Johnson et al., 1993; Duckett & Ryden, 1994) examined the performance of students in professional programmes, nursing, business and education, respectively. Duckett and Ryden (1994) noted that nursing practice includes an important moral dimension, so evaluated the relationship between DIT score and clinical performance. However, they did not did not describe a moral component within either the curriculum or the clinical performance measure used. They reported a significant correlation between DIT score and a measure of clinical performance \( r = 0.58, P < 0.001 \) among a group of 48 nursing students in their junior and senior years; DIT score accounted for 34% of the variance in the clinical performance of these students. This provides indirect evidence that the use of postconventional moral reasoning contributes to clinical performance, but the nature of the link is not yet clear. Johnson et al. investigated the relationship between students’ facility with business writing and moral judgement. Using a sample of 72 juniors and seniors, they found that GPA was the best predictor of DIT score, accounting for 70% of the variance. Similarly, they also found a significant relationship between students’ grades on a series of writing assignments that were scored for writing mechanics, completeness, tone and design. Whether earning high scores on these aspects of good writing constitutes moral behaviour is arguable. The noteworthy similarity between these two studies is that moral judgement was significantly related to performance variables within their respective disciplines.

Two studies examined performance of preservice teachers. Taylor et al. (1985) also investigated the general relationship between maturity in intellectual, moral and social domains and the quality of preservice teachers’ interpretations of child behaviours. They found no significant relationships between DIT scores and the quality of students’ interpretations. Similarly, Cartwright and Simpson (1990) found no significant relationship between DIT scores \( M = 36 \) and teaching effectiveness of 53 student teachers.
Bredemeier and Shields (1984) examined the moral dimension of students’ behaviours in the context of collegiate basketball games. They studied athletic aggression among 46 collegiate basketball players. Based on Bredemeier’s (1975) article, they defined aggression as “the initiation of an attack with the intent to injure within a sport context” (Bredemeier & Shields, 1984, p. 141). The behavioural-dependent variables they used were the average number of fouls committed per minute per season game for each player, coaches’ ratings of players’ aggressiveness (extremely low to extremely high aggression level) and their rankings of the aggressiveness of all players in relation to their teammates. They found a significant inverse relationship between P-score and athletic aggression on all three measures and several variations by gender (the correlation for number of fouls was significant only for the males, and the correlation with the ratings was much higher for women; women had significantly higher P-scores than their male counterparts). The definition of aggressiveness used in this study makes clear the link to moral development; this study shows that for both genders, moral judgement is related inversely to sports aggression.

Hubbs-Tait and Garmon (1995) noted that decisions regarding whether to engage in sexual activity are often included in the domain of moral dilemmas, and that the moral dimensions of these decisions have been heightened since the spread of the AIDS virus. They hypothesised that level of moral reasoning would mediate the relationship between AIDS knowledge and sexual behaviour. Using a sample of 103 single college students, they found a non-significant relationship between AIDS knowledge and sexual behaviours, but a significant negative correlation between DIT score and condom use. Next, they performed a median split on the DIT scores, creating a “high” group ($n = 45; \text{DIT scores } > 26.7$) and a low group ($n = 58; \text{scores } < 26.7$). Among those in the high group, a significant negative correlation was found between knowledge and degree of risk: as knowledge of AIDS increased, risky sexual behaviours decreased. This pattern was not found among the low group, where risky sexual behaviours did not decrease as AIDS knowledge increased. Given the overall low DIT scores for this sample, these relationships are especially provocative, suggesting that moral judgement does mediate behaviour even among college students with relatively low DIT scores (see [1]).

Hay (1983) explored becoming a conscientious objector (CO) as an act that is experienced by some as a response to a moral dilemma (whether to serve in the military). He administered the DIT to a sample of young adults who had responded to compulsory draft registration by registering as COs in order to determine whether their scores were comparable to those of non-COs. He found that at each educational level (high school, college, and graduate school), COs scored about 10 points higher than Rest’s norms (see [1]); the mean DIT score for the college students ($n = 101$) in his sample was 53. Hay also found that those who based their objections on a personal moral code (as opposed to religious training or belief, a criterion used by draft boards evaluating these appeals) had higher moral reasoning scores.

Lupfer et al. (1987) reported the results of four experiments examining student decision-making. Two of these took place in the context of a simulated trial in which the students (volunteers from an introductory psychology class) served as jurors.
The dependent behavioural variable was the number of guilty votes cast in the simulation. After taking the DIT, three types of six-person juries were formed: those with “high” P-scores (>29), (those with low P-scores (<21) and those with a mixture of scores. (The median score was low 25; see [1]) High-Pjuries cast significantly fewer guilty votes than low-P or mixed juries. In the mixed juries, those with higher leadership scores tended to have higher P scores ($r = 0.47, P < 0.001$), and these leaders successfully promoted their preference for acquittal; this phenomenon was not observed in the other jury types. Two problems make it difficult to interpret these findings: first, the DIT scores were quite low, even among the “high” group; second, the authors do not explain why acquittal is morally preferable. If it is, then the findings strongly suggest that higher moral reasoning is associated with moral behaviour; if not, then other explanations should be offered to explain this observed relationship.

This section has summarised studies that link collegiate contexts to moral judgement development. Based on these studies, it is clear that a wide variety of factors can foster moral development, which is important as there are few distinct experiences that are shared by most college students. As Pascarella and Terenzini (in preparation) concluded, “The key role of college in fostering postconventional moral reasoning may therefore lie in providing a range of intellectual, cultural, and social experiences from which a range of different students might potentially benefit” (p. 15). What is noteworthy is that most studies here examined the effects of students’ experiences in classroom contexts, and that comparable attention has not been given to structured co-curricular experiences (e.g. participation on disciplinary boards or in other student leadership positions) that have potential to affect moral development.

The Relationship between Moral Judgement and Other Domains of Development

This review has included only attributes that are particularly germane to college students or the college experience. As a result, we did not include attributes that simply reported relationships with other variables from, for example, a personality perspective instead of a college student or college outcomes perspective. Even with these parameters in place, a full analysis of the shared relationships between moral judgement and other college-related domains of development is beyond the scope of this paper. However, we have compiled a list of studies that shows the breadth of topics that have been researched comparing DIT scores with those from other domains. These studies address the following relationships: political orientation ($n = 12$); ego/identity development ($n = 11$), religious identification/religious attitude/spirituality ($n = 12$); social attitude/issues ($n = 9$); cognitive development ($n = 8$), academic achievement ($n = 8$), locus of control ($n = 7$); aptitude/intelligence ($n = 6$); values/ethics ($n = 5$); perspective-taking ($n = 4$); emotion ($n = 2$); tolerance/diversity ($n = 2$), learning/analytical style ($n = 2$); and conflict resolution ($n = 1$). (This list is available upon request from the authors.)
Suggestions for Future Research

These findings not only provide a summary of past research, but insights that can be used to inform and direct future research. One important conclusion is that dramatic gains in moral judgement are associated with collegiate participation, even after controlling for age and entering level of moral judgement. This provides an important springboard for future research; we next suggest the types of studies that will further our understanding of moral judgement development in the college years.

First, what accounts for differences in gains by institutional type? Most of the studies noted here only alert us to the existence of these differences, not to their causes; the Whiteley (1982) and Mentkowski et al. (2000) studies offer important exceptions, in that each of these documented the effects of specific, intentional educational activities. In the other studies, it is not clear whether obtained institutional differences are a result of other factors, such as: student self-selection (are those more open to thinking about moral dilemmas inclined to select a liberal arts college?); size (e.g. access to opportunities for direct involvement, smaller classes with more discussion); clarity of expectations regarding the role of values in students’ lives; opportunities to practice reasoning about moral issues; or a liberal arts curriculum that stresses examining many issues from multiple perspectives. Discerning these influences not only across but within types of institutions is an important topic for future research.

Secondly, how can the finding that a remarkably diverse set of interventions appears to promote moral judgement development be explained? What characteristics do these have in common? Is it making the moral dimension explicit? Using experiential learning strategies? Grounding discussions in personally meaningful topics? We urge researchers interested in the question of what makes a moral intervention effective to consult the extensive literature on this topic that has used school-aged samples, and (as discussed above) to employ rigorous research designs. Insights might also be gleaned from carefully examining those collegiate contexts that do not appear to promote moral development (e.g. fraternities and accounting courses). What are the characteristics of those collegiate subcultures? Are findings that students in these settings score lower than average an artefact of self-selection? A culture that gives higher value to non-moral attributes? Other factors? Investigating questions such as these could also contribute to our understanding of how collegiate contexts can support the development of moral judgement.

Thirdly, research in other domains has documented the powerful effects of student involvement in co-curricular activities (Kuh et al., 1991; Pascarella & Terenzini, 1991), yet this context is minimally reflected in this review. The effects on moral judgement of student involvement in community-outreach activities, campus judicial systems, leadership positions, peer-conflict mediation programmes and intergroup dialogue programmes remains largely unexplored.

Fourthly, how do educators (faculty, student organisation advisers, housing professionals, etc.) envision their responsibility for promoting moral judgement development? How do they enact this vision in their interactions with students?

Fifthly, there is a great deal of variability among college students, including but
not limited to race and ethnicity. A clearer understanding of student characteristics and student subcultures would inform our understanding of the kinds of experiences that enhance or inhibit moral development, and whether these reflect subgroup differences. Further, recent research on the power of campus climate variables (Hurtado et al., 1999) offers some rich concepts and examples to guide other research.

Sixthly, given the complexity of both the construct and of institutional contexts, larger-scale studies utilising more sophisticated statistical techniques are needed to untangle factors that lead to the development of moral judgement. Few of the studies reviewed here partialled out the effects of covariates (such as SES, which is correlated strongly with ethnicity and educational level); none used causal modelling or hierarchical linear modeling to discern and weigh effects on moral development.

In addition to identifying these broad topics for future research, we now offer a few suggestions for researchers interested in conducting research using the DIT. First, the domain of morality is much broader than the moral judgement component that the DIT is designed to measure, and we urge researchers to be cognisant of its purpose and intended use. For example, some researchers described it as a measure of “morality”, “ethics”, “values” or “social reasoning”. Although these are related domains, they are conceptually distinct from moral judgement and arguably inconsistent with the purpose of the DIT. Others used DIT scores to yield a single stage score as a proxy measure for Kohlberg’s moral interview; for a review of the conceptual and methodological problems with this approach, see Thoma (1994).

Secondly, we urge researchers who are interested in the development of moral judgement to clarify their rationale and underlying assumptions regarding their selection of variables to investigate. Showing the conceptual linkages they are testing offers much more fertile ground for theory-building than do studies that leave the reader guessing about these proposed relationships.

Thirdly, well-designed intervention studies are very helpful in identifying useful ways to promote moral judgement development. We urge researchers to carefully evaluate their proposed studies against established criteria for intervention studies, along the lines suggested above.

Finally, several of the studies reported here failed to report relevant data (e.g. P-scores for subgroups, sample sizes and mean scores) that were essential for meaningful interpretation. Others mislabelled groups after breaking down the data (e.g. labelling as “principled moral reasoners” those with DIT scores that were above the median for the sample, but still low compared to college students in general). These practices are fraught with potential for misunderstanding. We also urge researchers to carefully report their findings so that readers can arrive independently at their own conclusions about the appropriateness of the conclusions drawn. We believe that if more researchers follow these basic guidelines, then the next generation of DIT research will provide a clearer picture and better inform our understanding of the moral development of college students.
Conclusion

The DIT has been used in an impressive array of studies to examine various aspects of the lives and experiences of college students. Educators concerned with broad issues of character development in college students will be reassured to learn that development in moral reasoning does tend to improve during the college years. Further, specific collegiate contexts (liberal arts colleges, certain types of educational experiences) are also associated with growth in moral judgement, and with the production of moral behaviour. These findings offer a significant foundation for the future work of both researchers and educators.

This array of studies demonstrates that college students’ ability to base their moral judgements in postconventional moral reasoning does not simply unfold ontogenetically by chronological maturation alone. Rather, it is important to remember that development occurs in context, and that colleges offer excellent contexts to stimulate moral reasoning. (Whether educators take full advantage of these contexts is another matter.) Even a gross indicator of context such as institutional type suggests that students at some types of liberal arts colleges have higher potential to develop moral reasoning than they do at other types of colleges or universities.

This review has focused only on the DIT as a measure of moral judgement. The availability and demonstrated validity of this measure has clearly enabled the production of the plethora of studies summarised here. In the same spirit in which Jim Rest designed the DIT, we also urge researchers to continue to address theoretical questions, offer conceptual refinements, suggest alternative ways to measure both moral judgement and moral development and subject them to the scrutiny of sustained scholarly inquiry. Doing so would be a testament to the scholarly values that Rest inspired in the next generation of scholars whose own work is so richly represented in this volume.

Correspondence: Dr Patricia M. King, Center for the Study of Higher and Postsecondary Education, University of Michigan, 2117 School of Education Building, 610 E. University Avenue, Ann Arbor, MI 48109-1259, USA; E-mail: patking@umich.edu and Matthew J. Mayhew, Center for the Study of Higher and Postsecondary Education, University of Michigan, 2117 School of Education Building, 610 E. University Avenue, Ann Arbor, MI 48109-1259, USA; E-mail: mmayhew@umich.edu

NOTES

[1] According to Rest (1994), the mean DIT P-scores by educational levels are as follows: junior high students = 21.9; senior high students = 31.8; college students = 42.3; moral philosophy and political science graduate students = 65.2.
REFERENCES


A study of moral judgement, using the Defining Issues Test, for three ethnic groups at Bacone College, unpublished doctoral dissertation, Oklahoma State University, OK.


Enacting Diverse Learning Environments: improving the climate for racial/ethnic diversity in higher education, ASHE-ERIC Higher Education Report, 26(8), Washington, DC.


Developing Reflective Judgement: understanding and promoting intellectual growth and critical thinking in adolescents and adults (San Francisco, CA, Jossey-Bass).


Involving Colleges: successful approaches to fostering student learning and development outside the classrooms (San Francisco, CA, Jossey-Bass).


Incorporating the development of


