Journal of Moral Education, Vol. 31, No. 3, 2002

Carfax Publishing
Taylor & Francis Group

The Defining Issues Test and the Four Component Model: contributions to professional education

MURIEL J. BEBEAU
University of Minnesota, USA

ABSTRACT *This article reviews studies examining the effect of professional education on ethical development. Most studies limit assessment to the measurement of moral judgement, observing that moral judgement plateaus during professional school unless an ethics intervention is present. Whereas interventions influence the shift to postconventional reasoning (the DIT P score), a more illuminating picture of change may emerge if researchers examined DIT profiles. More importantly, limiting assessment to measures of moral judgement ignores important aspects of moral functioning suggested by the Four Component Model. Assessment methods have been validated for sensitivity, reasoning, role concept and ethical implementation that could be adapted to provide individuals in a particular profession with a more complete picture of abilities needed for real-life professional practice.*

Professional practice is predominantly a moral enterprise. Consequently, professional schools are concerned with the ethical development of their students. In the last two decades there has been a flurry of activity in designing and implementing ethics instruction, which previously had been a part of the "hidden curriculum". Because the resurgence of interest originated with moral philosophy rather than moral psychology, the contributions of moral psychology have not been a predominant force guiding curriculum design and assessment development. On the other hand, when psychologists have collaborated with philosophers in the design and assessment of instruction, both theory and practice have benefited. After providing a brief history of the ethics education movement, the contributions of the Defining Issues Test (DIT) to understanding moral judgement development in the professions are reviewed. With the exception of two studies in Nursing, the review is limited to studies in post-baccalaureate professional programs. Following a summary of findings within each of five professions, lessons learned from reflecting across professions suggest methods, as well as benefits, from incorporating the DIT into an ethics curriculum. Further, suggestions are given for the use of new DIT indices to give a finer-grained analysis of intervention effects. The second section

ISSN 0305-7240 print; ISSN 1465-3877 online/02/030271-25

DOI: 10.1080/0305724022000008115

describes the development of moral reasoning in the larger context suggested by Rest's (1982) Four Component Model (FCM) and summarises research and development that has extended our understanding of the processes that contribute to effective moral functioning.

I. A Brief History of Ethics Instruction in the Professions

Most educators working in professional schools today recognise the importance of ethics instruction to professional education. Many of the arguments that were launched in the late 1970s and 1980s—that ethics could not or should not be taught, that professional school was too late, or that faculty were ill equipped to teach it—seem to have fallen by the wayside. In the mid-1970s, however, when the ethics movement first took hold in the United States, ethicists had to argue for the importance of instruction that focused on ethical reasoning (see Hastings Center Report, 1980). Even though Harvard President Derek Bok (1976) argued that ethics instruction needed to attend to objectives besides reasoning, the emphasis of most courses was on some sort of moral problem-solving that involved attention to working out which of two conflicting values should take precedence in a given situation [1]. In the early days, the predominant method for resolving moral issues (Beauchamp & Childress, 1979) was the application of principles of health-care ethics (autonomy, non-maleficence, beneficence and distributive justice) to the resolution of cases. And, when assessment of ethical decision-making did occur, the preferred model was the analysis of written responses to cases (Howe, 1982). In this context, some educators used measures such as the DIT but, in the main, philosophers and ethicists who were pushing for ethics instruction in the professions argued against the use of measures such as Rest's DIT (e.g. Caplan) on the basis that the test was grounded in a Rawlsian/Kantian concept of justice that was biased against other moral frameworks that could and should be applied to resolve tough problems in the professions. Interestingly, despite the criticism of Kohlberg's theory, many ethics educators were relying on a principled approach to resolving moral issues. It was not until the early 1990s that the challenges to "principlism" (Clouser & Gert, 1990) resulted in a rethinking of principled approaches to resolving moral issues (see Beauchamp & Childress, 1994). Just as the challenges to principlism encouraged a rethinking of approaches to resolving ethical issues in the professions, such challenges influenced a data-based rethinking of what was measured by the DIT (Rest *et al.*, 1999). Evidence supports the DIT as a measure of postconventional moral thinking, rather than a measure of principled thinking. See Thoma (2002, this issue) for a brief discussion of the distinction.

II. Contributions of DIT Studies

King and Mayhew (2002, this issue) observed that many of the studies that used the DIT with college students were not intentionally designed to measure the moral judgement of college students or to study college as an educational context for development, with some exceptions (Rogers, 2002, this issue). Rather, college

students were used as a convenience sample, or as a proxy for young adults. In contrast, articles that report the use of the DIT with post-baccalaureate professional students were used exclusively to measure moral judgement development of students and to assess the curriculum as a context for development. Essentially three kinds of studies are reported: (1) those that compare subgroups of students, faculty or practising professionals within a profession and perhaps across institutions or even across professions; (2) those that contrast beginning students with graduates from the same programme—usually to establish some kind of baseline for judging development prior to implementing an educational intervention; and (3) those that use the DIT in a pre–post-evaluation study in an effort to document the effectiveness of an educational intervention. One of the challenges for professional educators is to determine whether an educational intervention is effective, since it is rarely possible to randomly assign students to experimental and control groups. Consequently, one of the frequent criticisms of this body of work and any attempted claim of educational effectiveness is the lack of contemporaneous control groups. In fact, Ruth Fisher, in a presentation to the Institute of Medicine Committee on Assessing Integrity in the Research Environment, argued that the greatest need for research on research ethics was funding for control group studies to assess the effectiveness of various interventions. However, experience in professional education shows that such studies are neither practically possible, nor likely to be approved by an institutional review board (Dubois, 2002).

Toward Understanding Moral Judgement Development in the Professions

This section summarises findings related to moral judgement from 33 studies (approximately 6600 respondents) from five professions. Most studies used the DIT, but studies using other measures (e.g. Sociomoral Reflection Measure (SRM) and Moral Judgement Interview (MJI) are cited as findings are generally consistent with DIT studies, thereby adding support for the conclusions. Four general questions guided the review: (1) does professional education promote moral judgement development? (2) Does the addition of ethics instruction promote reasoning development? Are some ethics courses more effective than others? How much curriculum time needs to be devoted to ethics instruction to influence reasoning development? Are changes in reasoning sustained? (3) Are there differences in moral judgement development among subgroups within a profession? For example, are surgeons, who confront life and death moral issues on a daily basis, better able to reason about complex issues than family practice physicians? Or, are attending physicians in teaching hospitals better able than residents and medical students to reason about moral issues? (4) Is moral judgement linked to professional performance?

Curriculum Effects

Medicine. With respect to whether medical education promotes personal communication, reasoning development, six studies (Daniels & Baker, 1979; George, 1997, 2002; Husted, 1978; Self *et al.*, 1991a; Sheehan *et al.*, 1981) report moral

judgement scores at various intervals during medical school and residency training in either cross-sectional or repeated measures designs. Although Sheehan *et al.* (1981) reported a moderate (4.3 points), but statistically significant, P score gain ($p < 0.02$) from first to third year for 52 medical students, no statistically significant differences were observed in the other five studies, suggesting that neither medical school curriculum nor residency training influences postconventional moral thinking. For example, Husted (1978) compared 488 first- and third-year medical students in a cross-sectional design, reporting no cohort differences. Daniels and Baker (1979) compared 60 medical students at semester intervals in a repeated measures design, and George (1997, 2002) compared 264 residents at various intervals in a repeated measures design. Neither researcher observed statistically significant change. Similarly, Self *et al.* (1991a) observed no significant change on the MJI for 20 volunteers who took the MJI at entry and exit from medical school.

Veterinary medicine. Six studies by Self and colleagues report curriculum effects for veterinary medicine. Four of these investigate whether the educational process (without ethics instruction) promotes reasoning development. One of the studies (Self *et al.*, 1993) reported a significant change ($p < 0.05$) from first to fourth year for 57 students (47% of the class) who were recruited to complete the SRM. In contrast, Self *et al.* (1996) reported no change for 54 of 111 students recruited to take the DIT before and after the first quarter of the first year. Similarly, no change on DIT P score was observed for 68 students (56.7% of the class) who completed the DIT at the beginning and end of the four-year curriculum. Further, Self, and colleagues (1991a) did not observe significant change for 20 students (16% of the class) who participated in MJI interviews at the beginning and again at the end of the four-year programme.

Law. Conversations with law school educators and law students indicate that the courses in professional responsibility required of all law schools following Watergate are not highly regarded and are seldom inspiring—as focus tends to be on learning a code of professionalism, rather than engaging in reflective thinking. A study by Willging and Dunn (1981) tested 41 University of Toledo law students before and after a 30-hour required course in professionalism. Whereas P scores were moderately high (52.22) at pretest, no significant change on the P score was observed following the course. In 1995, Hartwell conducted a series of studies with University of San Diego law students to test the effect on DIT P scores of student-centred small-group discussion of moral problems in law. To evaluate the effects of curriculum without ethics discussions, six studies were conducted. One examined whether a traditionally taught professionalism course (i.e. lectures and discussion of the code) would affect change ($n = 38$); five studies examined whether small group learning present in courses in negotiations and interviewing ($n = 150$), rather than moral discourse, was responsible for change. Neither the traditionally taught profes-

sionalism course nor the small-group learning courses showed change on DIT scores from pretest to posttest.

Dentistry. Two studies in dentistry investigate curricular effects prior to implementing an ethics intervention. Bebeau and Thoma (1994) reported mean DIT P scores for two classes of end-of-year juniors (47.6, SD = 13.3) with three groups of incoming freshmen (46.7, SD = 12.4). No statistically significant differences were observed between classes or among cohorts. Distinct from studies in medicine and veterinary medicine, all students completed the assessments as a curriculum requirement. In a similar attempt, Chaves (2000) compared mean P scores for three cohorts of entering freshmen ($n = 261$) with three cohorts of seniors ($n = 140$), prior to implementing a problem-based ethics course. Participation in the assessment was required for freshmen, but not for seniors. Consequently, only about 60% of the senior cohorts participated, compared with full participation by the first-year groups. Chaves (2002; personal communication) indicated that mean scores ranged from 40.3 (12.75) to 41.8 (13.06) for freshmen and from 38.4 (12.45) to 40.2 (16.08) for seniors. No significant differences were observed between classes or among cohorts.

Intervention Effects

Medicine. Of the 16 studies reporting development of moral reasoning, seven investigate various intervention effects. Five (Givner & Hynes, 1983; Self *et al.*, 1989, 1992; Self *et al.*, 1993; Self & Olivarez, 1996) reported significant pre/post change with gains ranging from 4.5 ($p < 0.05$) to 10 points ($p < 0.00025$). Two studies (Self & Olivarez, 1996; Self & Baldwin, 1998) reported the stability of gains over time, and one study (Self *et al.*, 1998) investigated the minimum number of contact hours required to achieve a significant gain of 4.3 points ($p < 0.006$), concluding a minimum of 20–29 hours was needed. Except for the study by Self *et al.* (1989), who compared a lecture course with a case study format against a contemporaneous control group that received no ethics course, all studies relied on students to volunteer to complete pre- and post-tests. For example, in a study of the amount of small-group instruction required to achieve change (Self *et al.*, 1998), 76% of 960 students (classes of 1991–1998) completed both tests. Volunteer rates of about 50% characterised the other three studies.

Veterinary medicine. In a study to assess the effects of ethics instruction, 105 of 128 veterinary medical students provided pre- and post-test data before and after a required ethics course that consisted of seven hours of lecture and eight hours of case-study discussion. Whereas there were no overall differences from pre-test (42.2) to post-test (42.7), female students showed a 4.3 point increase (43.7–46.7) which was statistically significant ($p < 0.02$). Because the 38 males regressed from pre-test to post-test (39.4–36.7), the mean gender difference at post-test (46.7 vs. 36.7) was significant ($p < 0.0005$).

Law. Hartwell (1995) tested the effect on DIT P scores of student-centred small-group discussion of moral problems in law. Advertised as an experimental course, 24 students tested 51.8 at pre-test and 61.8 at post-test. The 10-point gain was significant ($p < 0.05$). To replicate the effect, he conducted the course again, this time with 30 students who tested 44.45 at pre-test and 58.37 at post-test. The 14-point gain was significant ($p < 0.01$). A third replication of the course dropped the "experimental billing". Again, with 24 students enrolled, a significant 10.2-point gain ($p < 0.05$) was observed from pre-test (40) to post-test (50.2). Although Hartwell did not report standard deviations, Landsman and McNeel (2002) estimated that the effect size for the three cohorts of law students ($n = 78$) would be in the range of 0.77–0.97, assuming that the SD fell within the range of 12–15.

Nursing. Two intervention studies from Nursing are included in this analysis, as the studies are similar to the studies reported by Self and colleagues in that students were asked, but not required, to complete both pre- and post-test assessments. In contrast to the studies by Self and colleagues (response rates ranging from 48 to 95%), participation was unusually high in the studies by Duckett and Ryden (1994) and Duckett *et al.* (1997), with response rates ranging from 85 to 100% across the cohorts. Because the studies were conducted in the context of a well-designed and comprehensive ethics curriculum (Ryden *et al.*, 1989), with multiple replications (involving 368 students, and all but one of the eight cohorts showed a significant gain from entry to graduation), the studies are of interest. Because standard deviations are reported, it was possible to calculate effect sizes, which ranged from 0.16 to 0.65, with an average of 0.45. Yet, comparison of the nursing studies with studies in post-baccalaureate professions are limited, because studies have not controlled for the effect of maturation, which is evident in college student samples. Duckett *et al.* (1992) confirmed the effect of college in a review and secondary analysis of DIT data of 16 studies published in the nursing literature. The authors conclude that "moral reasoning of nurses, like that of other [college student] groups, tends to increase with formal education. Nurses' scores are usually comparable to, and sometimes higher than, scores of their academic peers" (p. 324).

Dentistry. In the 1994 study, Bebeau and Thoma concluded that change in moral judgement could be attributed to the Minnesota ethics curriculum. Their claim was based on a comparison of effect sizes for each of eight cohorts that completed the curriculum, with effect sizes drawn from a meta-analysis of intervention studies. The lack of mean differences from the cross-sectional comparison was used as a basis for arguing that the curriculum, prior to implementing the ethics intervention, did not promote reasoning. Thus, the effect was more likely attributable to the intervention. To extend the earlier findings, Bebeau (2001) analysed intervention effect sizes for seven subsequent cohorts, reported findings across studies, and investigated change using the new DIT indices. The average entering Minnesota dental student scores at 46 (with cohorts ranging from 42 to 49 across the 15 classes tested). The average graduate selects postconventional arguments 51% of the time (with cohorts ranging

from 47 to 55). Effect sizes vary across classes, with a range of 0.12–0.78, with an average of 0.43. Classifying students' change scores into categories defined by the standard error of measurement indicated that 44% of the 1229 students who participated in the curriculum made moderate to highly significant gains, 40% showed no change and 16% regressed on the P score. The observation of what appeared to be regression prompted a more thorough analysis of intervention effects. Bebeau *et al.* (2002) observed change not reflected in the P score: either a substantial decrease in the Personal Interests schema score or an increase in the Maintaining Norms schema score. For students that regressed, regression tended to be associated with a shift from transitional to consolidated status. Use of the new indices suggested that analysing change in terms of the P score was masking changes that are significant: acquisition of new thinking, rejection of simplistic thinking and being able to distinguish among moral arguments.

Subgroup Differences

Of the 33 studies reviewed, 18 addressed questions related to differences within a profession. These questions may seem less important to the seasoned moral judgement researcher, but professional educators ask them frequently. Many of these studies simply report mean differences. When standard deviations are reported, effect sizes are calculated.

Faculty and practitioners vs. residents and students. One question asked frequently when professional schools decide to implement an ethics curriculum is whether the reasoning of faculty and practitioners is more developed than that of students. Candee *et al.* (1982) observed significantly higher DIT raw scores for 186 faculty and practitioners (29.7, SD = 8.71) compared with 266 paediatric residents (27.45, SD = 8.71). The effect size is 0.25. George (2002, personal communication) observed significantly lower DIT N2 scores for internal medicine residents (37.2) compared with residents in family practice (45.5) and surgery (46.0). Husted's (1978) study suggested DIT differences for faculty from high prestige (54.2) vs. lower prestige (47.6) institutions, and George's studies (1997; 2002) of residents and faculty suggest that faculty have higher scores than residents. In veterinary medicine, however, data for 304 veterinary medical students suggests mean DIT scores ranging from 42 to 46. In contrast, data from a survey of 350 practising veterinarians (Self *et al.*, 1988) suggest that DIT scores, at least for 37% of practitioners who voluntarily returned the survey, are lower (33.8) than scores for veterinary students and than college graduates. For physician groups, Baldwin and Bunch (2000) did not find subgroup differences for two samples ($n = 80$) of orthopaedic surgeons (one sample had been used to establish the relationship between DIT scores and malpractice claims).

Cohort differences. In a study investigating the potential for an admissions interview, Benor *et al.* (1984) report a P score of 50.08 (17.0) for 38 applicants admitted to a community-based Israeli medical school based on an interview, compared with a P

score of 39.36 (12.8) for 161 rejected applicants. No such differences were observed when decisions were based solely on admission tests. The P score of 44 applicants admitted to a traditional Israeli medical school was 39.47 (12.4), compared with P scores of 40.02 (13.2) for 135 rejected applicants. In a study of medical students, Givner and Hynes (1983) observed significantly higher DIT pretest P scores (50.25 vs. 45.75) for 51 students who kept their promise to repeat the DIT post-test following a humanities course, compared to 57 who did not. Promise keepers also showed significant gains on the post-test. With respect to law students, the review of intervention studies suggests considerable variation among student cohorts, with scores ranging from 40 to 52. Two samples of practising attorneys, a small sample of 18 lawyers practising health law (Earnest & Bebeau, 2000) and a larger, although not representative, sample of 169 practising lawyers from a southern state (Scofield, 1997), suggests that practising lawyers and law students' scores are similar, with mean scores in the mid- to upper 40s.

Regional differences. Several observations suggest that researchers are likely to find regional differences. Summing across 15 cohorts (Bebeau & Thoma, 1994; Bebeau, 2001), the average P score of 45.64 (12.82) for entering Minnesota students is considerably higher than the mean (38.6, S.D. = 12.9) for 88 entering freshmen at UCLA (Coulter, 2002, personal communication). Similar to Minnesota students, UCLA students clearly took the test seriously, as none of the students failed the consistency check using "New check". The observed differences may be a function of the cultural mix, which is considerably different for the two schools. In contrast to Minnesota samples, which are predominantly Caucasian (over 90%), only 42% of the UCLA sample were Caucasian and 54% Asian. In contrast to the UCLA sample, the Indiana University sample is similar to the Minnesota sample, yet the mean scores for three cohorts were closer to the college student norm than to the graduate student norm. Finally, the mean for 31 of 43 University of West Virginia graduates with valid DIT scores (Nash *et al.*, 1982) was slightly lower (34.8) than the norms for college graduates, similar to the mean for 144 Indiana University graduates (39.1, SD = 12.82) and considerably lower than the average mean (51.33) of Minnesota graduates.

Gender. Consistent with the gender differences reported by Thoma (1986) that favour women, researchers have observed statistically significant differences between men and women in medicine (e.g. Self & Baldwin, 1998), veterinary medicine (Self *et al.*, 1995) and law (Landsman & McNeel, 2000). Such differences would not be noteworthy were it not for the magnitude of the observed differences. Thoma (1986) observed an overall effect size of 0.213 for 18 samples ($n = 1539$) of college students and 0.279 for four samples ($n = 370$) of graduate students. The 95% confidence intervals for the graduate student samples were 0.07–0.50 compared to 0.11–0.32 for college students. In a study designed to evaluate the stability of change following a first-year ethics course, Self & Baldwin (1998) compared gender differences for 488 of 851 Texas A & M medical students from 1991 to 1995. Sixty percent of the student body is male, 40% female. Of the 57% of the student body that provided

cross-sectional data, 61% were male. Statistically significant gender differences ranged from 5.5 to 8.4 points across the four classes with P scores for males ranging from 49.8 (12.6) to 51.1 (16.6) and P scores for females ranging from 55.3 (13.4) to 58.7 (12.9). Because standard deviations were reported for each cohort, it was possible to calculate effect sizes: 0.38 for 1st-year, 0.40 for 2nd-year, 0.62 for 3rd-year and 0.53 for 4th-year students. Landsman and McNeel (2002) tested 170 (81%) of the entering law students at the University of Minnesota. They reported a mean for 76 females of 53.1 (12.39) and 47 (16.10) for 88 males. The mean difference was statistically significant ($p < 0.007$), with an effect size of 0.42. The veterinary medicine intervention study (Self *et al.*, 1995) described above raises questions about the differential effects of instruction for males and females. In the above studies the pooled standard deviation ranged from 13.57 to 15.41. Estimating the effect size based on a pooled standard deviation of 15.41, the effect size is 0.64; if 13.57, the effect size is 0.74. In either case, the effect size falls outside the confidence intervals reported by Thoma. The differences between men and women in these samples is truly larger than those reported for the more general samples, with each of the estimates above the mean effect size of the graduate samples, and three of the six estimates outside the confidence intervals of earlier studies.

Culture. It may be tempting to attribute the mean P score differences reported by Husted (1978) for 46 US trained residents (57.2) with the mean of 32.3 for 58 foreign-trained residents to cultural differences. However, Duckett and colleagues (1992) warn of possible problems with confounding variables. In their reanalysis of a study by Nokes, American-trained nurses with higher DIT scores also scored significantly higher on tests of reading comprehension than foreign-born nurses. Because the DIT is heavily dependent on reading skills, its use with subjects whose first language is not English needs to be considered carefully before making any inferences about cultural differences.

Links to Performance

Medicine. Studies linking moral judgement with clinical performance have been reported repeatedly (Sheehan *et al.*, 1980, 1985), and are not summarised here. However, a study by Candee *et al.* (1982) is summarised, as the significance of that study with respect to the context for ethical decision making has not been discussed. The study takes on added value in view of the recent interest in promoting academic integrity and research integrity. Candee *et al.* (1982) reported that the content of a physician's attitude is related to the structure of moral thought. Specifically, "residents whose moral reasoning was more developed were less aggressive in their treatment of neonatal defects, especially in cases where the family requested such limits, or where the quality of life after treatment was likely to preclude even a minimal degree of human or social interaction. Further, they showed greater variability among cases, suggesting a more complex or more discriminating approach to treatment" (p. 849). To control for the influence of the norms and culture

of the institution in which the resident practised, Candee examined the amount of variance accounted for (4%), after accounting for the influence of institutional type (high prestige vs. low prestige)—which accounted for 43% of the variance in level of activism. As the same relationship between reasoning and activism was not observed fully qualified practitioners, the study raises questions about the power of culture to shape consistency in judgement over time.

Law. Landsman and McNeel (2000) investigated the links between 170 entering law students' moral judgement development (mean P score = 49.6, SD = 14.86) and preferences for 12 law subdisciplines. Moral judgement was significantly higher when students more highly valued altruistic law disciplines. Whereas political view was the strongest predictor of perceived importance of public interest jobs (beta weights of 0.34), P score and sex were equally important in predicting such interest (beta weights of 0.22 and 0.21, respectively). Landsman and McNeel are following these students to see how practice preferences change as students move through the curriculum to the job market.

Nursing. Some of the strongest evidence for the relationship between moral judgement and clinical performance comes from the studies conducted at the University of Minnesota. Critical to the effort to establish such a relationship is the validation of a measure of clinical performance. Krichbaum and her Minnesota colleagues (Krichbaum *et al.*, 1994) validated a Clinical Evaluation Tool (CET) and, after establishing inter-rater reliability, asked faculty to rate students at the end of the three-year curriculum. Of the 85 students in the study, complete data were available for 48. After establishing the comparability of students eliminated because of incomplete data on one or more variables, forward stepwise multiple linear regression indicated that DIT P score and age at admission accounted for 46% of the variance in the mean CET score. DIT P scores, entered at Step 1, accounted for 34% of the variance; age, entered at Step 2, accounted for the additional 12%. Whereas all the cognitive measures (the DIT, ACT (American College Testing) and GPA (Grade Point Average)) had significant zero order correlations with clinical performance, only the DIT contributed significantly to predicting the CET scores in the multiple regression. Krichbaum and colleagues concluded that the study confirmed "the long-held belief that both cognitive maturity and moral maturity are critical elements of nursing practice". Although the generalisability of the finding is limited to the Minnesota programme, the findings are congruent with the work of Sheehan *et al.* (1980). Further, the presence of a well-validated measure of clinical performance should enable other researchers to extend these findings. In the eight years since Krichbaum established the strong relationship between clinical performance and reasoning, Minnesota nursing faculty (Duckett, 2002, personal communication) have had difficulty in securing valid CET ratings for students. In part, this is due to the increasing reliance on community-based clinical courses and rotations

with preceptors and temporary faculty who are less experienced in teaching and evaluating students.

Dentistry. Nash *et al.* (1982) attempted to establish the relationship between clinical performance and moral reasoning observed in medicine and later in nursing. No relationship was observed. However, the authors commented that the measure of clinical performance focused heavily on the technical dimensions of competence, rather than on aspects of the doctor–patient relationship. Subsequently, Meetz *et al.* (1988) validated a measure of clinical performance for assessing the relationship between reasoning and clinical performance—broadly defined. Results (Bebeau, 1994) indicate that the relationship is not linear as observed by Sheehan *et al.* (1980), but the datasets have one thing in common: low scores on reasoning exclude the possibility of high scores on measures of clinical performance. As in the nursing curriculum, efforts to replicate the finding have been met with frustration. First, while it is possible to get reliable data on clinical performance by using Meetz's measure, in the current configuration of clinical experience faculty work only with small groups of students and fail to have a broad perspective that enables reliable ranking of students. Alternative methods such as a productivity index (Bebeau, 2001) are very effective in identifying highly effective students, but because students must achieve a minimum index to graduate, the data tends to produce highly skewed distributions. The search for behavioural indicators is important not only for replicating the relationship between reasoning and behaviour, but also for other components of ethical development.

Lessons from Curriculum and Intervention Studies

There are obvious limits to generalising from these curriculum studies to professional education in general. For example, all the studies use convenience samples and most, except for the studies in dentistry, relied on volunteers to provide estimates of curricular and intervention effects. Were it not for the consistency across studies and across professions, it would be hard to suggest that these studies represent a reliable effect. The pooled findings from 10 crosssectional studies ($n = 1934$ students) and from 10 repeated measures studies ($n = 547$ students and residents) that show no effect when compared with two studies ($n = 106$) showing modest change, supports these conclusions. Although one might not want to go so far as to claim, as did Self and Baldwin (1998), that professional school inhibits growth in reasoning, available evidence suggests that the curriculum does not promote change, at least not for the volunteers who provided data. Whether volunteers are different from non-volunteers is a question of interest. Findings from the cross-sectional study involving 554 dental students is consistent with studies using volunteers.

Studies using the DIT in the professions have extended our understanding of what constitutes an effective intervention. There follow some lessons learned from the review of studies cited earlier.

What Works?

Whereas dilemma discussion is the technique most often recommended for facilitating development (Schlaefli *et al.*, 1985), other strategies appear to be as effective. For example, Self *et al.* (1993) observed significant effects for a film course that was comparable to the effects for their lecture and case-study method. Bebeau's (1994) curriculum, although it provides only about 12 hours in actual dilemma discussion over a four-year programme, accompanies that discussion with requirements for writing well-reasoned arguments that apply criteria for judging the quality of arguments that have been presented in advance of the small-group discussion. Hartwell's student-centred moral discourse seems to provide the largest gains, effect sizes estimated at 0.77–0.97. Compared with the average effect size (0.45) calculated from the data provided by the nursing studies, or the average effect sizes (0.43) reported by the dentistry studies, the effect achieved with the technique used by Hartwell is large, similar to the 0.80 effect size (McNeel, 1994) attributed to college.

Providing Feedback on the Pre- and Post-test

Except for the nursing studies, none of the researchers indicate whether students are given information about their pre- or post-test scores. The way the information is presented is as important as giving feedback. In the dental curriculum, students received a personal letter reiterating what was said when introducing the test before informing them of their individual scores. At the end of the curriculum, a similar letter details results of the DIT, as well as progress on other measures of ethical development. The letter stressed that students are lifelong learners, and that the test information was designed to help them set goals for their future continuing education, and help them appreciate the extent to which they might rely on their own judgement when encountering novel problems. As the curriculum focused on collaborative decision making and sharing of perspectives, we hoped we had developed that habit of consultation with others when challenging problems arise.

Including the DIT as an Outcome Measure

If researchers want to use the DIT to study the effects of curriculum or the effects of an intervention, we think there are sufficient reasons to include the measure as a regular part of the curriculum. Asking students to complete the test for curricular purposes voluntarily, other than their own empowerment, seems unjustified. Not only are response rates usually insufficient to provide useful information about the impact of instruction, but testing students without involving them in the questions of interest seems disrespectful and unethical. Further, there appear to be sufficient reasons to include the DIT as a measure of outcome. When the test is appropriately presented and constructive feedback is given, the test can serve a useful curricular purpose.

III. Contribution of the Four Component Model

As is apparent from studies reviewed in the previous section, the major focus of research on ethical development in the professions has been on moral judgement. This is unfortunate, as the DIT assesses only one dimension of professional performance. Other dimensions clearly impact on professional performance. Even at the inception of the current ethics movement, educators (Bok, 1976) were arguing for the importance of abilities like those described by the Four Component Model (FCM). In fact, the correspondence between the abilities recommended by Bok (1976) and the Hastings Center Report (1980) is remarkable, given the differences in their disciplinary origins. Yet the FCM—grounded in a review of psychological research—extends substantially philosophers' conceptions of ethical sensitivity and ethical implementation. The FCM identifies at least four integrated abilities as necessary conditions for effective moral functioning. This section summarises efforts to design methods and measures for assessing abilities suggested by the FCM. After describing the measures and findings with respect to each, the explanatory power of the model is discussed.

Ethical Sensitivity

Ethical sensitivity involves the ability to interpret the reactions and feelings of others. It involves being aware of alternative courses of action, knowing cause–consequence chains of events in the environment and how each could affect the parties concerned. As such, it involves empathy and role-taking skills. For individuals being socialised to professional practice, ethical sensitivity involves the ability to see things from the perspective of other individuals and groups (including other cultural and socio-economic groups), and more abstractly, from legal, institutional and national perspectives. Thus, it includes knowing the regulations, codes and norms of one's profession, and recognising when they apply. In professional settings, the focus is on ethical sensitivity, rather than the more general "moral sensitivity", to signal the distinctive expectations of the profession that derive from the norms and codes that govern professional practice.

Assessment

Performance-based methods for assessing ethical sensitivity were first developed in dentistry (Bebeau *et al.*, 1985), and the most extensive work on construct validity has been conducted with the Dental Ethical Sensitivity Test (DEST Forms A and B). See Rest *et al.* (1986) and Bebeau (1994, 2001) for summaries of the validation studies. The general strategies for ethical sensitivity assessment have been applied in other professional settings: counsellor education (Volker, 1984), computer users (Liebowitz, 1990), undergraduate education (Mentkowski & Loacker, 1985; McNeel, 1994), geriatric dentistry (Ernest, 1990), social work (Fleck-Henderson, 1994), journalism (Lind, 1997) and school personnel (Brabeck *et al.*, 2000). An ethical sensitivity test places students in real-life situations, where they witness an

interaction either on videotape or audiotape. The interaction replicates professional interactions and provides clues to a professional ethical dilemma. Studies assessing the ethical sensitivity of both professionals in training and professionals in practice (Bebeau *et al.*, 1985; Bebeau & Brabeck, 1987; Fleck-Henderson, 1994; Bebeau, 2001) indicate considerable variability among professionals, in sensitivity to the embedded ethical issues. Thus, completion of professional training does not ensure development of sensitivity to professional issues. However, studies also show that ethical sensitivity can be improved with instruction (Mentkowski & Loacker, 1985; Bebeau & Brabeck, 1987; Leibowitz, 1990; Sirin *et al.*, submitted). Further, studies show (Bebeau *et al.*, 1985; Bebeau & Brabeck, 1987; Brabeck *et al.*, 2000) that ethical sensitivity is distinct from the ability to reason (Rest's second component) about what ought to be done in the situation. Consequently, one cannot assume that education that focuses on ethical reasoning will influence to the interpretive process.

Moral Reasoning and Judgement

Moral judgement is a critical element of professional ethical development, and tests such as the DIT have a place in assessing lifespan moral judgement development. In fact, Strike (1982) argued that acquisition of sophisticated and abstract principles of moral reasoning, as measured by the DIT, may be a necessary condition, but insufficient for effective functioning. The question (for educators) is often whether to teach specifically to the codes (as in law) or policy manuals (as in research ethics), or to teach concepts particular to a discipline—informed consent, intellectual property, conflict of interest, etc. Strike refers to the latter as "intermediate level ethical concepts", as they lie in an intermediate zone between the more general principles (e.g. autonomy, justice, beneficence) described by philosophers, and the more prescriptive directives often included in codes of conduct. Studies in dentistry highlight the importance of assessing profession-specific aspects of moral judgement.

Assessment

Written ethical arguments. Perhaps the most familiar approach to assessing reasoning in professional ethics courses is the analysis of written arguments typically conducted by faculty with background in philosophy (Howe, 1982). These techniques have been applied in dentistry (Bebeau, 1994) and nursing (McAlpine *et al.*, 1997) and the studies have demonstrated that essays can be reliably assessed and that instruction is effective in promoting the ability to develop well-written essays that meet criteria, if the criteria are specified in advance of instruction. Such methods lack practicality for assessing competence as a function of an institution's efforts to promote reasoning about dilemmas in professional ethics. However, assessment of written essays is a particularly effective way to promote learning, especially when accompanied by clearly stated criteria, frequent opportunities for practice, and feedback (Bebeau, 1994a). The methods have been applied to research ethics in a collaborative project with Indiana University's Poynter Center. As its title

implies, *Moral Reasoning in Scientific Research: cases and materials* (Bebeau *et al.*, 1995) is designed to facilitate improvement in moral reasoning, as well as assessment of such improvement. The booklet, available on the Poynter Center's website (http://www.indiana.edu/ ~poynter/mr-main.html), features case studies, facilitator notes, and a handout for students that details criteria for judging the adequacy of moral arguments.

Testing intermediate-level ethical concepts. The Dental Ethical Reasoning and Judgment Test (DERJT) is a first effort to test application of context-specific concepts (taught in ethics courses) to real cases (Bebeau & Thoma, 1999). The test is similar to the DIT, in that cases are presented followed by lists of action choices and justifications. When taking the test a respondent rates each action or justification, then selects the two best and two worst action choices, and the three best and two worst justifications. Scores are determined by calculating the proportion of times a respondent selects action choices and justifications consistent with "expert judgement". In validation studies, clear expert/novice differences are observed. Further, scores for students, practitioners, and referrals appear to be normally distributed. In a study comparing Minnesota graduates' responses to familiar vs. unfamiliar problems presented on the test, it appears that a good grasp of postconventional moral schemas is a necessary condition for transfer to new problems.

Moral Motivation and Identity Formation

Component three acknowledges that individuals have legitimate concerns that may not be compatible with the moral choice. Career pressures, established relationships, and idiosyncratic personal concerns compete for the professional's attention. Lapses of professional behaviour can often be attributed to low priority placed on moral considerations, even when the moral choice is well understood. Blasi (1985) notes that people differ in how deeply moral notions penetrate their self-understanding, and in the kinds of moral considerations that are judged constitutive of the self. Understanding that one is responsible provides the bridge between knowing the right thing to do and doing it. In professional contexts, moral motivation and commitment has to do with the importance given to professional values in competition with other values. Deficiencies in motivation are apparent when personal values such as self-actualisation or protecting one's self or organisation replace concerns for doing what is right.

Identity formation is seen (Kegan, 1982; Blasi, 1985) as a lifelong developmental process. Whether a professional identity has formed by the time a person enters a profession is of considerable interest. Recently, using Kegan's measure of identity formation, Forsythe and colleagues (2002) interviewed West Point cadets at various stages of the educational process. They reported that at least 30% of West Point graduates have not achieved key transitions in identity formation that would enable them to have the broad, internalised understanding of codes of ethics or the commitment to professional standards. Without internalised standards, codes and

professional regulations may be perceived as guides for behaviour, and such individuals are likely to conform to the guides, in order to garner rewards and avoid negative consequences. Forsythe and colleagues (2002) concluded that unless the broad educational environment promotes identity development toward a shared perspective on professionalism, professional development programmes would not be successful in instilling desired values in less mature pre-professionals. Studies of role concept development in dentistry (Bebeau, 1994) support Forsythe's observations. Students entering the profession are not able to articulate the key concepts of professionalism. Even after instruction and practice, some students are unable to do so, suggesting that the conceptual frameworks for a professional identity are not a part of initial self-understanding, and must be revisited frequently during professional education.

Identities vary by level of maturity and sophistication. Recent work in research integrity documents the need for more formal efforts to socialise researchers to professional expectations and values. For example, Braxton and Baird (2001) argue that if future scientists are expected to participate in the deterrence, detection and sanctioning of scientific wrongdoing, doctoral study must socialise researchers to the role of self-regulation. The need for such socialisation is highlighted by Anderson's (2001) study of doctoral students' conceptions of science and its norms. Based on in-depth interviews, she concludes that students are not learning the normative aspects of the research enterprise by "as much osmotic group socialization as many faculty assume". The need for "deliberative approaches to normative socialization find support in the vagueness with which students conceptualize the norms that underlie academic research."

Assessment

Two assessment methods can be used to evaluate role concept development. One is to ask students (at various stages of their education) to write a short essay entitled: "What does it mean to become a dentist, or physician, attorney, nurse, or scientist?" Such an essay can be critiqued based on the extent to which the norms and values that undergird the profession are described. Each educator can develop his or her own criteria for assessment of the essay, based upon the instruction provided on the topic. (See Bebeau, 1994, for an example.) A second method is the use of a measure of role concept, perhaps patterned after measures developed in dentistry, medicine and law.

The Professional Role Orientation Inventory (PROI) (Bebeau *et al.*, 1993) assesses commitment to privilege professional values over personal values. A series of studies attest to the construct validity and test–retest reliability (Thoma *et al.*, 1998); additional studies are in progress. Bebeau (2001) reported statistically significant change from pretest to post-test for five cohorts of Minnesota dental students who took the test as freshmen and again as seniors. Cross-sectional studies of differences between pre- and post-test scores for a comparable dental programme suggests that ethics instruction accounts for change.

The Professional Decisions and Values Test (Rezler *et al.*, 1992) was designed

for lawyers and physicians to assess action tendencies and the underlying values in situations with ethical problems. Patterned after the Defining Issues Test, the test consists of 10 case vignettes, followed by three alternative actions and seven reasons to explain the chosen action. Findings support the construct validity of the test, test–retest reliability is stable over time for action choices, but not for values.

Ethical Implementation

Fundamental to responsible conduct in any profession is ability to perform with integrity the complex tasks of the discipline. The fourth component in the FCM attends to the importance of character to effective and responsible practice. A practitioner may be ethically sensitive, may make good ethical judgements and place high priority on professional values; but if the practitioner wilts under pressure, is easily distracted or discouraged, or is weak-willed, then moral failure occurs because of a deficiency in character and competence. Professional education assists individuals in understanding the fundamentals of their discipline, gaining depth in the details of a particular sub-area, and obtaining practical experience through clinical practice, residency training and advanced speciality training. Fischer and Zigmond (1998), speaking in the context of research ethics, point out that an essential dimension of such educational programmes is often neglected. They argue for the development of a set of general professional skills that, when performed badly, reflect negatively on the individual's integrity. Assessing essential ethical implementation skills requires performance-based assessment.

Assessment

Objective measures have not been devised to measure competence in implementing effective action plans. Further, it is hard to imagine designing anything but performance-based assessments of the broad range of skills required for effective, responsible professional practice. Objective Structured Clinical Examinations (OSCEs) have been designed in medicine that approximate the kind of integrated performance required for assessing "character and competence". In the dental programme, students are directed to prepare: (1) an interpretation of facts presented in a case; (2) an action plan; and (3) verbatim dialogue to illustrate how the action plan can be implemented. Peers and clinicians, using checklists validated for each case, judge responses. Responses to a series of such cases serve as a proxy for judging competence in ethical implementation. An alternative to observations of interactions is to collect examples of professional performance for evaluations by faculty and students similar to the portfolios Gilmer (1995) has students develop for her courses in research ethics.

The Explanatory Power of the Four Component Model

Over the last 20 years, I have made numerous presentations on the FCM to groups

of educators and practising professionals. When moral failings are conceptualised as deficits in abilities that can be remediated through education, educators tend to be more hopeful about the potential for ethics education. When practitioners are referred by a licensing board for ethics instruction, they often assume that they are viewed as people who lack integrity. Helping them to see that they have deficiencies in one or more abilities related to ethical decision-making has restorative power. They can learn to see things as others see them, to reason more carefully, to rethink their priorities and to change communication habits that undermine their goals. Further, careful exploration of their viewpoints can help them to see where their thinking is flawed, and often why their actions lead to disciplinary action on the part of their colleagues.

The model has intuitive appeal, and educators come away with a more refined view of the goals for courses in professionalism. Yet, key to helping an individual identify personal shortcomings are profession-specific measures, such as the DEST, the DERJT, the PROI and the analysis of verbatim dialogues offered in response to real-life professional problems. Because criteria for judgement are grounded in professional expertise, and because norms are available to help the individual compare his/her ability to others, the assessment is viewed as authentic, and professionals are willing to submit to the learning process. Further, when dealing with mature adults with considerable intellectual ability, learning often moves quickly, and gains following 20–30 hours of personalised instruction (augmented by reading, writing and role-playing) are marked. Further, one consistent observation, in addition to a deficiency in either sensitivity, reasoning or implementation, is the difficulty 34 of the 35 referrals had in articulating the expectations of the professional. In contrast, for moral exemplars (Bebeau, 2002), such responsibilities appear to be at the forefront of their thinking and a driving force that guides action. After targeted instruction, directed toward role concept development and remediation of one or more other deficiencies, we observed measurable improvements in performance, coupled with documented changes in the behaviours that gave rise to the disciplinary action. Further, to date, there have been no cases of recidivism [2]. Examining case studies bolsters the understanding of the connection between the components and behaviour, and provides direction for education.

Independence of the Components

In a summary of the evidence related to the independence of the components, Bebeau (2001) observed that studies typically report low to very low correlations between ethical sensitivity and moral judgement, but correlations among the other components have varied from very low to an occasional moderate correlation. In some cases there appears to be some overlap between subscales of some measures. Conclusions to date suggest that measures of the components are assessing abilities that are distinct from one another.

IV. Summary and Conclusions

In the professions, the DIT has been used primarily to study curricular effects. Unlike college education, professional school curricula seems not to promote reasoning development—as measured by the P score—unless there is an ethics component that involves students in the discussion of ethical issues. Whereas a variety of ethics interventions seem effective, student-centred moral discourse described by Hartwell (1995) produced the largest effects, and ought to be explored by other professions. Once change has been achieved for a group, it seems to be stable over time. However, the studies reviewed report change only in terms of the P score, and then in terms of the mean pre- to post-test change. Reporting findings only in terms of mean gains on the P score appears to be masking other kinds of change (Bebeau *et al.*, 2002) that are significant to professional education. One effect is the acquisition of new thinking represented by increases in the P score. A second effect is systematic rejection of simplistic thinking represented by decreases in the Personal Interest schema score. A third is the ability to clearly distinguish among competing arguments—the move from transitional to consolidated status. From a practical educational point of view, each kind of developmental progress is desirable: gaining more sophisticated moral thinking, becoming clearer about what ideas to reject for their simplistic and biased solutions, and being able to distinguish among moral arguments and their theoretical origins. Use of the newer indices now included in the DIT scoring service should enable educators to provide a more illuminating picture of change. If the DIT is to be used to study curricular effects, we recommend that educators introduce it as an outcome measure for the curriculum, rather than presenting it as a voluntary activity. The DIT is an exceptionally well-validated and reliable measure, and the development of reflective moral judgement is an aim for professional education. It may be the case, as Haidt (2001) argues, that individuals make intuitive judgements that are not informed by reason, but such unreflective decision making is not consistent with competent professional practice. The ability to develop well-reasoned arguments for recurrent and emerging professional problems is an important goal for professional education. Useful as the DIT can be for assessing a student's general approach to resolving moral issues, it should not be the only measure of moral judgement used in an ethics curriculum. Courses focus typically on the acquisition of profession-specific concepts (e.g. informed consent), and essays or norm-referenced tests such as the DERJT are needed to assess the acquisition of a well-thought-out rationale for addressing authentic professional problems.

For educators wanting to develop an assessment programme to accompany a curriculum in professional ethics, the DIT is certainly a starting point. Most adults find the measure interesting to take, and if introduced well and if careful feedback is given, the test can be a useful adjunct to instruction. However, neglecting other abilities suggested by the FCM leaves professionals at risk. Individuals need to know whether they can reliably interpret ethical issues, articulate the norms, values, laws and codes that govern professional practice and implement defensible action plans effectively and efficiently. Instruction without measures to help students see their

strengths and shortcomings, and to compare them to peers as well as to seasoned and exemplary colleagues, is unlikely to promote competence. Further, professional school educators need to take seriously their responsibility to design authentic assessments to demonstrate that instruction makes a measurable difference in abilities that relate to everyday moral functioning.

One reaction of readers may be to question the amount of progress made in the more than 20 years since the FCM was first proposed. Indeed, as this review shows, only a handful of researchers have designed measures to assess other dimensions of competence. Whereas considerable resources are expended to develop and teach professional ethics (Rest & Narvaez, 1994 estimated that some 10 000 courses are taught), less attention has been given to assessment of learning. In large part, this is the result of inadequate funding for studies to design and validate appropriate measures. Because individual teachers or even individual institutions are unlikely to be able to mount the kind of research and development plan needed to design and validate measures that assess the important outcomes of professional ethics education, an effort that crosses institutional and professional boundaries is needed. The recent concerns about integrity in research may be the catalyst required for mounting such an effort [3]. A second reason has to do with the origin of the professional ethics education movement. Courses tend to be taught by educators with background in philosophy or team taught by ethicists and practitioners. Learning outcomes may be specified, but are neither theoretically grounded nor tied to assessment. Courses tend to focus on moral theories and their application to cases. Validated assessment methods are seldom used. Our view (Bebeau *et al.*, 1999) is that rather than arguing that ethics outcomes resist quantitative or qualitative analysis, greater attention needs to be directed to systematic research and development using well-validated measures and a theory of development that focuses attention on identifying processes as they contribute to moral action.

Future Study

Future study should be directed at methods and measures for assessing components of morality other than moral judgement. Two areas seem particularly promising. (1) Ethical sensitivity research has not progressed very far because the assessment process is relatively expensive, requiring transcription of a semi-structured interview and scoring by trained raters. Recently, however, Brabeck and Sirin (2001) produced a computerised version of the Racial Ethical Sensitivity Test (REST-CD) intending to make their test more efficient. Sirin *et al.* (submitted) concluded that the more efficient assessment process provides a reliable and valid measure of ethical sensitivity to instances of racial and gender intolerance. The modified assessment strategy needs to be tested in other professional settings. (2) Work on identity formation and role concept development would advance our understanding on professional ethical development. Promising as the work by Forsythe *et al.* (2002) is, the assessment process, like the ethical sensitivity process, is expensive and time-consuming. Easier to administer objective measures are needed. Similarly, further work is required on role concept measures to assess acquisition of norms and

values specific to a profession. Measures such as the PROI seem to differentiate among groups expected to differ on professional values, and the methods could be applied to other professions.

One aspect of socialisation that has not been studied adequately is the influence of moral climate and culture of professional development. As suggested by the study by Candee and colleagues (1982), the structure of individuals' moral judgement does influence professional decision-making, especially for people at entry level within the profession. However, the norms and culture of the institution seem to have a powerful effect, shaping consistency in decision making across individuals. The power of moral climate was first investigated in moral psychology (Higgins *et al.*, 1984), has been investigated in organisational development, and has recently become the focus of attention for academic integrity and research integrity (Institute of Medicine, 2002). By extension, climate and culture will need to be addressed in professional education. The Minnesota approach to morality research, as exemplified by the dental and nursing ethics programmes, has shifted the focus from development of moral judgement, as measured by the DIT, to the development and assessment of ethical sensitivity, identity formation and the abilities related to ethical implementation. Findings from this review provide substantial evidence that shifting attention to the development and assessment of other abilities, using profession-specific measures derived from the FCM, enhances our understanding of moral functioning and meaningfully contributes to student learning. Professional ethics education could benefit from making a similar shift.

Correspondence: Dr Muriel J. Bebeau, Professor, Department of Preventive Sciences, School of Dentistry, University of Minnesota, 515 Delaware Street SE, Minneapolis, MN 55455, USA; E-mail: bebea001@umn.edu

NOTES

[1] A number of ethicists (e.g. Rule & Veach, 1993) published frameworks for the discussion of cases, which were essentially a series of steps to go through when confronted with a professional problem.

[2] Practitioners have been referred for breaches in the state dental practice act, not for problems with impulse control, mental illness or substance abuse.

[3] The problem related to funding for measurement development was recently taken up by an Institute of Medicine Committee charged with making recommendations to the Office of Research Integrity for "Assessing Integrity in Research Environments". As a member of the committee, I (MJB) contributed a chapter on assessment and collaborated with other committee members on a chapter on education. Both rely on the Four-Component Model for theoretical grounding for curriculum and assessment recommendations. The report (Institute of Medicine, 2002) includes recommendations related to funding.

REFERENCES

ANDERSON, M. (2001) What would get you in trouble: doctoral students' conceptions of science and its norms, proceedings of the ORI Conference on Research on Research Integrity. Available on the ORI website: http://www-personal.umich.edu/~nsteneck/rcri/index.html

BALDWIN, D.C. JR. & BUNCH, W.H. (2000) Moral reasoning, professionalism, and the teaching of ethics to orthopedic surgeons, *Clinical Orthopedics and Related Research*, 378, pp. 97–103.

BEAUCHAMP, T.L. & CHILDRESS, J.F. (1979, 1994) *Principles of Biomedical Ethics*, 4th edn (New York, Oxford University Press).

BEBEAU, M.J. (1994) Influencing the moral dimensions of dental practice, in: J. REST & D. NARVAEZ (Eds) *Moral Development in the Professions: psychology and applied ethics*, pp. 121–146 (New York, Lawrence Erlbaum Associates).

BEBEAU, M.J. (2001) Influencing the moral dimensions of professional practice: implications for teaching and assessing for research integrity, proceedings of the ORI Conference on Research on Research Integrity. Available on the ORI website: http://www-personal.umich.edu/~nsteneck/rcri/index.html

BEBEAU, M.J. (2002) *Identity formation and moral motivation*, paper presented in the session on Moral Exemplars in Oral Health Care for PEDNET at the annual meeting of the American Dental Education Association, San Diego, CA, March 2–6.

BEBEAU, M.J., BORN, D.O. & OZAR, D.T. (1993) The development of a Professional Role Orientation Inventory, *Journal of the American College of Dentists*, 60, pp. 27–33.

BEBEAU, M.J. & BRABECK, M.M. (1987) Integrating care and justice issues in professional moral education: a gender perspective, *Journal of Moral Education*, 16, pp. 189–203.

BEBEAU, M.J., PIMPLE, K.D., MUSKAVITCH, K.M.T., BORDEN, S.L. & SMITH, D.L. (1995) *Moral Reasoning in Scientific Research: cases for teaching and assessment* (Bloomington, IN, Indiana University).

BEBEAU, M.J., REST, J.R. & NARVAEZ, D.F. (1999) Beyond the promise: a perspective for research in moral education, *Educational Researcher*, 28, pp. 18–26.

BEBEAU, M.J., REST, J.R. & YAMOOR, C.M. (1985) Measuring dental students' ethical sensitivity, *Journal of Dental Education*, 49, pp. 225–235.

BEBEAU, M.J., RODRIQUEZ, M. & MAEDA, Y. (2002) *Using DIT moral schema profiles to evaluate educational intervention effects*, paper presented at the annual meeting of the Association for Practical and Professional Ethics, Cincinnati, OH, February 28–March 3.

BEBEAU, M.J. & THOMA, S.J. (1994) The impact of a dental ethics curriculum on moral reasoning, *Journal of Dental Education*, 58, pp. 684–692.

BEBEAU, M.J. & THOMA, S.J. (1999) "Intermediate" concepts and the connection to moral education, *Educational Psychology Review*, 11, pp. 343–360.

BENOR, D.E., NOTZER, N., SHEEHAN, T.J. & NORMAN, G.R. (1984) Moral reasoning as a criterion for admission to medical school, *Medical Education*, 18, pp. 423–428.

BLASI, A. (1985) The moral personality: reflections for social science and education, in: M.W. BERKOWITZ & F. OSER (Eds) *Moral Education: theory and application*, pp. 433–443 (Hillsdale, NJ, Lawrence Erlbaum Associates).

BOK, D. (1976) Can ethics be taught? *Change Magazine*, October, pp. 26–30.

BRABECK, M.M., ROGERS, L.A., SIRIN, S. *et al.* (2000) Increasing ethical sensitivity to racial and gender intolerance in schools: development of the racial ethical sensitivity test, *Ethics and Behavior*, 10, pp. 119–137.

BRABECK, M.M. & SIRIN, S. (2001) *The Racial Ethical Sensitivity Test: computer disk version* (REST-CD), available from first author: M.M. Brabeck, Lynch School, Boston College, Chestnut Hill, MA 02467.

BRAXTON, J. & BAIRD, L. (2001) Preparation for professional self regulation, *Science and Engineering Ethics*, 7, pp. 593–614.

CANDEE, D., SHEEHAN, T.J., COOK, C.D., HUSTED, S.D. & BARGEN, M. (1982) Moral reasoning and decisions in dilemmas of neonatal care, *Pediatric Research*, 16, 846–885.

CHAVES, J.R. (2000) Assessing ethics and professionalism in dental education, *Journal of the Indiana Dental Association*, 79, pp. 16–21.

CLOUSER, K.D. & GERT, B. (1990) A critique of principlism, *Journal of Medicine and Philosophy*, 15, pp. 219–236.

DANIELS, M.H. & BAKER, G.L. (1979) *Assessing the moral development of medical students: an empirical*

study, proceedings of the 18th Annual Conference on Research in Medical Education, 18, pp. 87–92.

DUBOIS, J.M. (2002) When is informed consent appropriate in educational research? *Ethical and Regulatory Issues, IRB: Ethics & Human Research*, 24, pp. 1–8.

DUCKETT, L., ROWAN, M., RYDEN, M. *et al.* (1997) Progress in the moral reasoning of baccalaureate nursing students between program entry and exit, *Nursing Research*, 46, pp. 222–228.

DUCKETT, L., ROWAN-BOYER, M., RYDEN, M.B., CRISHAM, P., SAVIK, K. & REST, J.R. (1992) Challenging misperceptions about nurses' moral reasoning, *Nursing Research*, 41, pp. 324–331.

DUCKETT, L.J. & RYDEN, M.B. (1994) Education for ethical nursing practice, in: J.R. REST & D. NARVAEZ (Eds) *Moral Development in the Professions: psychology and applied ethics*, pp. 51–70 (Hillsdale, NJ, Lawrence Erlbaum Associates).

EARNEST, F. & BEBEAU, M.J. (2000) *Use of defining issues test in leadership education*, paper presented at the 26th Annual Conference of the Association for Moral Education, Glasgow, Scotland, July 6–11.

ERNEST, M. (1990) *Developing and testing cases and scoring criteria for assessing geriatric dental ethical sensitivity*, unpublished M.S. thesis, University of Minnesota, MN.

FISCHER, B.A. & ZIGMOND, M.J. (1998) Survival skills for graduate school and beyond, in: M. ANDERSON (Ed.) *The Experience of Being in Graduate School: an exploration* (New Directions in Education Series, no. 101), pp. 29–40 (San Francisco, CA, Jossey-Bass).

FLECK-HENDERSON, A. (1994) *Ethical sensitivity: a theoretical and empirical study*, unpublished Doctoral dissertation, Santa Barbara, CA, The Fielding Institute.

FORSYTHE, G.B., SNOOK, S., LEWIS, P. & BARTONE, P. (2002) Making sense of officership: developing a professional identity for 21st century army officers, in: D. SNIDER & G. WATKINS (Eds) *The Future of the Army Profession* (New York, McGraw Hill).

GEORGE, J.H. (1997) Moral development during surgical residency training, in: A.J.J.A. SCHERPBIER, C.P.M. VAN DER VLEUTEN, J.J. RETHANS & A.F.W. VAN DER STEEG (Eds) *Advances in Medical Education*, pp. 747–748 (Dordrecht, Kluwer Academic Publishers).

GILMER, P.J. (1995) Teaching science at the university level: what about the ethics? *Science and Engineering Ethics*, 1, pp. 173–180.

GIVNER, N. & HYNES, K. (1983) An investigation of change in medical students' thinking, *Medical Education*, 17, pp. 3–7.

HAIDT, J. (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment, *Psychological Review*, 108, pp. 814–834.

HASTINGS CENTER (1980) *The Teaching of Ethics in Higher Education* (New York, Hastings Center).

HARTWELL, S. (1995) Promoting moral development through experiential teaching, *Clinical Law Review*, 1, pp. 505–539.

HIGGINS, A., POWER, C. & KOHLBERG, L. (1984) The relationship of moral atmosphere to judgments of responsibility, in: W.M. KURTINES & J.L. GEWIRTZ (Eds) *Morality, Moral Behavior and Moral Development*, pp. 74–108 (New York, John Wiley & Sons).

HOWE, K. (1982) Evaluating philosophy teaching: assessing student mastery of philosophical objectives in nursing ethics, *Teaching Philosophy*, 5, pp. 11–22.

HUSTED, S.D. (1978) Assessment of moral reasoning in pediatric faculty, house officers and medical students, proceedings of the 17th Annual Conference on Research in Medical Education, 17, pp. 439–441.

INSTITUTE OF MEDICINE (2002) *Integrity in Scientific Research: creating an environment that promotes responsible conduct* (Washington, DC, National Academy Press).

KEGAN, R. (1982) *The Evolving Self: problems and process in human development* (Cambridge, MA, Harvard University Press).

KING, P. & MAYHEW, M. (2002). Moral judgement development in higher education: insights from the Defining Issues Test, *Journal of Moral Education*, 31, 3, 247–270.

KRICHBAUM, K., ROWAN M., DUCKETT, L., RYDEN M.B. & SAVIK, K. (1994) The clinical evaluation tool: a measure of the quality of clinical performance of baccalaureate nursing students, *Journal of Nursing Education*, 33, pp. 395–404.

LANDSMAN, M. & MCNEEL, S.P. (2000) *Moral judgment and preference for public interest law practice among beginning law students*, paper presented at the 26th Annual Conference of the Association for Moral Education, Glasgow, Scotland, July 6–11.

LANDSMAN, M. & MCNEEL, S.P. (2002) *Do women lose their idealism in law school? Gender, moral judgment and preference for [certain] public interest law practice*, unpublished paper, University of Minnesota, MN.

LEIBOWITZ, S. (1990) *Measuring change in sensitivity to ethical issues in computer use*, unpublished doctoral dissertation, Boston College, MA.

LIND, R. (1997) Ethical sensitivity in viewer evaluations of a TV news investigative report, *Human Communication Research*, 23, pp. 535–561.

MCALPINE, H., KRISTJANSON, L. & POROCH, D. (1997) Development and testing of the ethical reasoning tool (ERT): an instrument to measure the ethical reasoning of nurses, *Journal of Advanced Nursing*, 25, pp. 1151–1161.

MCNEEL, S.P. (1994) College teaching and student moral development, in: J. R. REST & D. NARVAEZ (Eds) *Moral Development in the Professions: psychology and applied ethics*, pp. 27–49 (Hillsdale, NJ, Lawrence Erlbaum Associates).

MEETZ, H.K., BEBEAU, M.J. & THOMA, S.J. (1988) The validity and reliability of a clinical performance rating scale, *Journal of Dental Education*, 52, pp. 290–297.

MENTKOWSKI, M. & LOACKER, G. (1985). Assessing and validating the outcomes of college, in: P.T. EWELL (Ed.) *Assessing Educational Outcomes. New directions for institutional research*, no. 47, pp. 47–64 (San Francisco, CA, Jossey-Bass).

NASH, P.J., NASH, D.A. & HUTTON, J.L. (1982) Moral reasoning and clinical performance of dental students, *Journal of Dental Research*, 61, Program and Abstracts, #268.

REST, J.R. (1982) A psychologist looks at the teaching of ethics, *The Hastings Center Report*, 12, pp. 29–36.

REST, J.R. & NARVÁEZ, D.F. (Eds) (1994) *Moral Development in the Professions: psychology and applied ethics* (Hillsdale, NJ, Lawrence Erlbaum Associates).

REST, J.R., BEBEAU, M.J. & VOLKER, J. (1986) An overview of the psychology of morality, in: J.R. REST (Ed.) *Moral Development, advances in research and theory*, pp. 1–39 (Boston, MA, Praeger).

REST, J., NARVAEZ, D., BEBEAU, M.J. & THOMA, S.J. (1999) *Postconventional Moral Thinking: a neo-Kohlbergian approach* (Mahwah, NJ, Lawrence Erlbaum Associates).

REZLER, A.G., SCHWARTZ, R.L., OBENSHAIN, S.S., LAMBERT, P., MCGIBSON, J. & BENNAHUM, D.A. (1992) Assessment of ethical decisions and values, *Medical Education*, 26, pp. 7–16.

ROGERS, G. (2002) Rethinking moral growth in college and beyond, *Journal of Moral Education* 31, p. 3.

RULE, J.R. & VEATCH, R.M. (1993) *Ethical Questions in Dentistry* (Chicago, IL, Quintessence).

RYDEN, M. DUCKETT, L., CRISHAM, P., CAPLAN, A. & SCHMITTZ, D. (1989) Multi-course sequential learning as a model for content integrations: ethics as a prototype, *Journal of Nursing Education*, 28, pp. 271–275.

SCHLAEFLI, A., REST, J. & THOMA, S. (1985) Does moral education improve moral judgment? A meta-analysis of intervention studies using the Defining Issues Test, *Review of Educational Research*, 55, pp. 319–352.

SCOFIELD, S.B. (1997) *Re-examination of the application of cognitive developmental theory to the study of ethics and socialization in the accounting profession*, PhD dissertation, Louisiana Tech. University, LA.

SELF, D.J. & BALDWIN, D.C. JR. (1998) Does medical education inhibit the development of moral reasoning in medical students? A cross-sectional study, *Academic Medicine*, 73, pp. S91–S93.

SELF, D.J. & BALDWIN, D.C. JR. & OLIVAREZ, M. (1993) Teaching medical ethics to first-year students by using film discussion to develop their moral reasoning, *Academic Medicine*, 68, pp. 383–385.

SELF, D.J., BALDWIN, D.C. JR. & WOLINSKY, F.D. (1992) Evaluation of teaching medical ethics by an assessment of moral reasoning, *Medical Education*, 26, pp. 178–184.

SELF, D.J., BALDWIN, D.C. JR., WOLINSKY, F.D. & SHADDUCK, J.A. (1993) Further exploration of the relationship between veterinary medical education and moral development, *Journal of Veterinary Medical Education*, 20, pp. 140–147.

SELF, D.J. & OLIVAREZ, M. (1996) Retention of moral reasoning skills over the four years of medical education, *Teaching and Learning in Medicine*, 8, pp. 195–199.

SELF, D.J., OLIVAREZ, M. & BALDWIN, D.C. JR. (1998) The amount of small-group case-study discussion needed to improve moral reasoning skills of medical students, *Academic Medicine*, 73, pp. 521–523.

SELF, D.J., OLIVAREZ, M., BALDWIN, D.C. JR. & SHADDUCK, J.A. (1996) Clarifying the relationship of veterinary medical education and moral development, *Journal of the American Medical Association*, 209, pp. 2002–2004.

SELF, D.J., PIERCE, A. & SHADDUCK, J.A. (1995) Description and evaluation of a course in veterinary ethics, *Journal of American Veterinary Medical Association*, 207, pp. 1550–1552.

SELF, D.J., SAFFORD, S.K. & SHELTON, G.C. (1988) Comparison of the general moral reasoning of small animal veterinarians vs. large animal veterinarians, *Journal of the American Veterinary Medical Association*, 193, pp. 1509–1512.

SELF, D.J., SCHRADER, D.E., BALDWIN, D.C. JR. & WOLINSKY, F.D. (1991a) The moral development of medical students: a pilot study of the possible influence of medical education, *Medical Education*, 27, pp. 26–34.

SELF, D.J., SCHRADER, D.E., BALDWIN, D.C. JR., ROOT, S.K., WOLINSKY, F.D. & SHADDUCK, J.A. (1991b) Study of the influence of veterinary medical education on the moral development of veterinary students, *Journal of American Veterinary Medical Association*, 198, pp. 782–787.

SELF, D.M., WOLINSKY, F.D. & BALDWIN, D.C. JR. (1989) The effect of teaching medical ethics on medical students' moral reasoning, *Academic Medicine*, 64, pp. 755–759.

SHEEHAN, T.J., CANDEE, D., WILLMS, J., DONNELLY, J. & HUSTED, S.D. (1985) Structural equation models of moral reasoning and physician performance, *Evaluation in the Health Professions*, pp. 379–400.

SHEEHAN, T.J., HUSTED, S.D. & CANDEE, D. (1981) *The development of moral judgment over three years on a group of medical students*, paper presented at the Annual Meeting of the American Education Research Association.

SHEEHAN, T.J., HUSTED, S.D., CANDEE, D., COOK, C.D. & BARGEN, M. (1980) Moral judgement as a predictor of clinical performance, *Evaluation in The Health Professions*, 8, pp. 379–400.

SIRIN, S., BRABECK, M.M., SATIANI, A. & ROGERS, L.A. (forthcoming) *Development of computerized racial ethical sensitivity test*.

STRIKE, K.A. (1982) *Educational Policy and the Just Society* (Chicago, IL, University of Illinois Press).

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

THOMA, S.J., BEBEAU, M.J. & BORN, D.O. (1998) Further analysis of the Professional Role Orientation Inventory, *Journal of Dental Research*, 77, p. 120.

THOMA, S. (2002) An overview of the Minnesota Approach to Research in moral development, *Journal of Moral Education*, 31, p. 3.

VOLKER, J.M. (1984) *Counseling experience, moral judgement, awareness of consequences and moral sensitivity in counseling practice*, unpublished doctoral dissertation, University of Minnesota, MN.

WILLGING, T.E. & DUNN, T.G. (1981) The moral development of the law student: theory and data on legal education, *Journal of Legal Education*, 31, pp. 306–358.