

## Chapter 13

---

---

### *Summary: What's Possible?*

James R. Rest  
Darcia Narváez  
*University of Minnesota*

The previous chapters describe many possibilities for research in applied ethics. Each chapter offers something different. If the various authors were to take each other's advice, what would a complete program of research look like? This chapter summarizes the possibilities.

Saying "complete program" does not imply that there is nothing else that can be done beyond what is suggested in this book. Rather, the chapters of this book suggest many kinds of studies that are technically possible at the current time. All of these suggestions, if applied to every profession, would comprise an ambitious program of research.

This chapter is organized around four themes: (a) describing moral reasoning with existing instruments; (b) discussing ideas for more effective educational interventions; (c) developing new measures for moral psychology; and (d) linking judgment to behavior.

### *DESCRIBING MORAL REASONING WITH EXISTING INSTRUMENTS*

---

One of the first studies that researchers do is to administer one of the standard moral judgment instruments (e.g., Moral Judgment Interview (MJI), Defining Issues Test (DIT), Sociomoral Reflection test, Haan Moral Judgment test) to various groups.

The choice of instrument depends on many considerations.<sup>1</sup> In the previous chapters we have discussed various reasons for choosing one test or another.

But what difference does a moral judgment score make, anyway? Following the argument in chapter 1, moral judgment scores represent the basic interpretative framework that people naturally and spontaneously bring to moral problem solving (the *default schema*, the bedrock of conceptions for making sense of moral dilemmas). A DIT *P* score below 50 means that the person is predominantly not conceptualizing moral problems the way moral philosophers conceptualize the problem (i.e., by determining what is morally right from the perspective of a society that balances the interests of its participants, optimizes the stake of each participant in supporting that society, and eliminates arbitrary advantages or influence). It means that low scoring students do not appreciate the ethics professor's reading list of wonderfully insightful articles by moral philosophers. Low scoring students see just a lot of words in brilliant arguments that wend their way through thickets of complications. For low scoring students, discussions of intermediate-level concepts (e.g., *informed consent*, *paternalistic deception*, *privileged confidentiality*—see chap. 1) do not find lodging in a bedrock of basic cognitive structure, but rather seem like superfluous solutions for problems neither foreseen nor recognized. For students with low moral judgment scores, it means that the principled solutions to ethical problems must be learned one at a time (as *special overrides*), largely by rote, since their default schemas do not provide a general perspective for anticipating principled solutions. These students have trouble extending principles beyond the cases specifically taught. They are baffled when ideals conflict. In real life, it means that people with low moral judgment scores are likely to oversimplify life situations, and although they might have good technical skills and generally good intentions, they are vulnerable to finding themselves involved in ethical problems over their heads.

Three kinds of studies using existing moral judgment instruments to describe levels of moral judgment predominate in the literature: (a) studies that compare one subgroup of professionals with another; (b) studies that compare students beginning a professional program with students finishing the program; and (c) studies that use existing tests for pretest and posttest evaluation of specific courses or interventions in moral education.

## Comparison of Subgroups of Professionals

Consider, first, comparison studies. In collecting moral judgment scores from, say, first year nurses or accountants or school teachers, often researchers compare the scores of these groups with other professional groups who have similar levels of

---

<sup>1</sup> The fact that the DIT is one of the easiest test to administer and score (being multiple-choice and computer-scored) should not be held against it. Despite its ease of use, there is no other program of research with other instruments that has produced clearer findings or more useful information about professional ethics. Although other instruments usually involve more pain, there is not inevitably more gain.

education. If the researcher finds that the target group scores lower than the comparison groups, then there is a concern that the target group is somehow deficient. This concern was raised by Ponemon and Gabhart for accounting students (chap. 6), by Chang for students going into public school teaching (chap. 4), by Self, Baldwin, and Olivarez for veterinary students (chap. 9), and by Bredemeier and Shields for athletes (chap. 10). On the other hand, in the nursing literature, Nokes (1989) mistakenly has stated that nurses have very low DIT scores (and she somehow concluded that nurses shouldn't be concerned with justice issues anyway); however, Duckett and Ryden (chap. 3; and Duckett et al. (1992) showed that nurses and nursing students are not deficient if the correct scores of the DIT are compared (Nokes confused raw scores with percentage scores—for instance, a raw score of 30 is a percentage score of 50).

Researchers have not only been interested in comparing one professional group with another, but also in comparisons within professions. For instance, Ponemon and Gabhart (chap. 6) found that senior members of accounting firms do not have higher scores than junior members, and that accountants from Canada have higher scores than those from the United States, McNeel (chap. 2) found that college seniors majoring in liberal arts have higher scores than seniors majoring in more vocational/careerist programs, Self and Baldwin (chap. 8) compared male and female medical students, and found that whenever there are differences, females are ahead of males on Kohlbergian tests of moral judgment, Duckett and Ryden (chap. 3) compared older students with younger students in nursing, and found the advantage going to the older students, Self and Baldwin described a study of applicants for medical school and reported that the DIT was related to admissions decisions—that is, applicants with higher DIT scores had a better chance of getting into medical school.<sup>2</sup> These, then, are some of the group comparisons that have been studied.

## Effects of Professional Education Programs

The second kind of study contrasts beginning students in a multiyear professional program with graduating students from the same program. In the summary of literature in chapter 1, it was reported that years of schooling may be correlated with moral judgment development. McNeel (chap. 2) reported a meta-analysis of 22 studies of college freshmen and college seniors. The effect size for 4 years of college for moral judgment is about .80. This is an effect size that puts moral judgment among the largest of any effects of college (see Pascarella & Terenzini, 1991, for a review of hundreds of studies). McNeel estimated that first year students in college average 35.7 on the DIT, and seniors average 46.4. In previous reports, only a composite of students in college in all years was given (42.3 in chap. 1). Thus, McNeel's data give us more precision.

---

<sup>2</sup> Self and Baldwin speak against using moral judgment scores as admission criteria; it is the exit score that is important, not the entry score. Since professional school can significantly improve moral judgment scores and make up this deficit, they argue that it should not be used to block entry.

Previous research (chap. 1) suggested that formal education has its effect upon moral judgment by virtue of continued intellectual stimulation. McNeel added to this picture by showing how dogmatism and indoctrination in a college environment are inhibitors of moral judgment development, and that liberal arts (in the classical sense of questioning, inquiring, and openness to evidence and argument) are enhancing. McNeel suggested that college programs that are either too careerist (too narrow a focus on the technicalities of initial job preparation) or too dogmatic (in closing off questioning and inquiry) inhibit growth in moral judgment. Therefore, colleges that promote conservative religious beliefs might inhibit growth if they are dogmatic and indoctrinating. But McNeel pointed out that being religious does not inhibit growth. He showed that in the case of Bethel College, a liberal arts approach is compatible with strongly held religious convictions, and that both foster moral judgment development.<sup>3</sup>

McNeel also found that on his campus the programs that emphasize careerism are the programs that slow moral judgment development. The emphasis on careerism may be due to the curriculum, to student characteristics, or both. What do the other chapters report about the effects of professional education on moral judgment?

In short, some of the chapters do report gains in moral judgment scores, and some do not. Self, Baldwin, and Olivarez (chap. 9) did not typically find gains over the program for veterinary students ("The experience of veterinary medical education appears to inhibit the increase in moral reasoning"). Similarly, Self and Baldwin (chap. 8) did not usually find gains in the moral reasoning of medical students over the course of their program. Ponemon and Gabhart (chap. 6) did not consistently find gains in accounting students; however, students in an accounting program within a liberal arts curriculum do seem to gain more. On the other hand, some multiyear professional programs did report gains in their students' moral judgment. Bebeau (chap. 7) reported gains after the incorporation of ethics components into the curriculum. Similarly, Duckett and Ryden (chap. 3) reported significant gains of students in their nursing program. In summary, some professional programs that emphasize practical/technical training show gains and others do not. We come back later to consider what makes the difference.

## Effects of Interventions

The third kind of study using standard instruments is the pre-post evaluation study of specific courses. Rest (1986, chap. 3) reviewed 56 such programs. Chapters in this volume by Sprinthall; McNeel; Duckett and Ryden; Self, Baldwin, and Olivarez; and Self and Baldwin also described evaluation studies. In these studies, generally the experimental or treatment groups statistically showed greater gains

---

<sup>3</sup> Dogmatism need not be limited to the *Right*. Dogmatism can also come from the *Left*, in the form that D'Souza (1991) called *political correctness*. It would be interesting to get data from some highly selective, high-priced liberal arts colleges in this country that have been swept with political correctness. Is it the case that they are actually inhibiting moral judgment development of their students?

than control or comparison groups. Let us now turn to characteristics of successful programs.

## MORAL EDUCATION INTERVENTIONS

The major suggestions for interventions will be discussed as follows: (a) integrating direct experience with reflection; (b) using the Four Component Model for designing instruction; (c) using the Multi-Course Sequential Learning model; and (d) incorporating didactic teaching of fundamentals of logic and philosophy.

### Direct Experience with Reflection

Sprinthall has been devising and developing educational interventions for over two decades (see chap. 5). He used pre–postevaluation data to tell him what educational interventions are working. His notion of *deliberate psychological education* fuses three elements: reading academic psychology, actively performing human service work, and a reflective seminar attempting to integrate the academic theory with the real-life experience. Sprinthall followed Dewey in asserting that immediate experience and active problem solving are crucial. A lecture course on theory does not tie down the concepts to something on the level of direct experience. In accord with reviews of the intervention literature (chap. 1), active practice in problem solving is important. But Dewey and Sprinthall went on to say that experience must be accompanied by creating a symbolic representation of the experience. The mere feeling that something significant happened at some time is not enough. In order to profit from the experience and have it as a resource to inform future encounters, a cognitive framework of understanding must be developed. Sprinthall, in his chapter, contended that experiential education without the opportunity for reflection (in the seminars) is not effective.

### Using the Four Component Model

Bebeau (chap. 7) is most explicit about using the Four Component Model as a guide to planning instruction and for evaluation of the effectiveness of that instruction. One of the concerns of this approach is the integration of affect and cognition. She describes the viewpoint that moral development involves several component processes and that moral education must be concerned with all of the components. Also, Duckett and Ryden (chap. 3) used the distinctions among components in their instruction—and subdivide Component IV into *ego strength* and *social skill* components. Similarly, Sprinthall's emphasis on roletaking, and McNeel's point about including empathy underscored the importance in integrating various elements into educational programs.

## Multi-Course Sequential Learning

Duckett and Ryden (chap. 3) described their approach to moral education in professional education in terms of Multi-Course Sequential Learning (MCSL). The basic idea here is that ethics is best not taught as a single-shot course, or as an incidental inclusion in courses (so integrated that you can't find it). Instead, the MCSL approach involves a series of units running throughout the program that address ethical issues as they naturally arise in the experience of the student. Their approach makes good curriculum sense, but requires an unusually high degree of deliberate curriculum explicitness and cooperation among faculty (in contrast to the more frequent Balkanization of the curriculum and hidden secrets of what goes on in different required courses). The approach described by Duckett and Ryden is a tribute to their faculty's cooperation and curriculum explicitness. The gains of their students attest to its effectiveness.

## Direct Teaching of Fundamentals

Penn's contribution to moral education was discussed in McNeel's chapter. Penn (1990, 1992a, 1992b) argued that certain logical and philosophical concepts are critical to the formulation of a principled perspective, and that these basics can be directly, didactically taught. Penn did not argue against student activity in problem solving; rather he argued that students need not invent everything anew from scratch. For instance, in teaching chemistry, students are not merely pointed to the lab and told to mix up something. Their practical, direct experience is first informed by some basic knowledge and guidance through the experiments. In the same way, Penn argued, moral education ought not be totally free-form grappling with moral controversies, but student learning can benefit from didactically teaching some basic logical and philosophical elements first. Then students are challenged with cases of moral problem solving. The theoretical reasonableness of Penn's argument is strengthened by empirical results. Penn showed some of the highest gains in students' DIT scores of any moral intervention. McNeel's chapter described some of the teaching methods and materials developed by Penn.

## Lack of Studies Demonstrating Consensus in Ethics

The four points just mentioned describe things that people are actually doing now to improve the effectiveness of educational interventions. But there is something else that needs mentioning, which is largely not being done. Recall from the preface that three assumptions about applied ethics were listed:

1. Assumption 1: Some ethical judgments are more justifiable than other judgments.
2. Assumption 2: There is some agreement among experts on moral judgments.
3. Assumption 3: Ethics courses affect students in some constructive ways.

Whereas attention has been given to Assumptions 1 and 3, too little attention has been given to Assumption 2. Ethics instructors throughout the country presumably evaluate students (in 10,000 courses annually, according to one estimate). In so doing, ethics instructors must be gathering some sample of student work or thinking, evaluating the work, and assigning grades or credits. What is lacking is demonstration that the evaluations by one instructor is comparable to the evaluation by other instructors—that there is such a thing as consensus among experts.

Bebeau (chap. 7) quotes students who state that ethical judgment is nothing more than idiosyncratic opinion, and describes the development of evaluation instruments that involve attaining consensus on criteria by experienced practitioners. For instance, setting the scoring criteria for her Dental Ethical Sensitivity Test (DEST) involves discussions and the consensus of practicing dentists, notably senior and highly respected practitioners, Fellows of the American College of Dentists. Hence, when scorers use these criteria and are checked for interjudge reliability, there is a basis for claiming that the evaluations represent more than just one person's idiosyncratic opinion. Further, use of the DEST by several schools around the country—sharing one set of consensually derived criteria—provides evidence that evaluation in ethics is not completely idiosyncratic. Moreover, it would be a good investment for ethics instructors at different institutions to exchange their evaluation tasks, criteria, and student sample work, and demonstrate that instructors can agree on the evaluation of student work. This would do much to document Assumption 2.

## *NEW INSTRUMENTS TO ASSESS MORALITY*

---

The interest in new instruments for assessing morality comes from two directions. The first interest is in devising more relevant, updated, profession-specific tests of moral judgment, (than the standard tests). The second interest stems from the realization that there are more aspects to morality than moral judgment as assessed in terms of Kohlbergian stages.

Consider first the issue of devising new tests of moral judgment (Component II), as Chang (chap. 4) and Westbrook (chap. 11) described. They wanted to use dilemmas more relevant to their specific subjects with the hope of strengthening the links of judgment with behavior. Let us consider how Chang and Westbrook faced four key decisions in devising new tests: Decision 1: the choice of a data-collection technique; Decision 2: establishing a basis for claiming that some judgments are better than others; Decision 3: devising a strategy for validation; and Decision 4: devising an index.

*Decision 1* concerns how data are to be collected. For instance, are data to be collected by interviews about moral dilemmas, yielding utterances that are matched with scoring criteria; by the dilemma-and-item format of the DIT, yielding ratings and rankings; by essays on current controversial topics; by responses to videotapes;

or by simulated role-playing situations? Both Chang and Westbrook followed the format of the DIT for their new instruments, but there is nothing magical about the DIT format. What is new about their instruments is that Chang used dilemmas specific to the teaching profession and Westbrook uses journalism dilemmas. Westbrook described the back-and-forth process with journalists that he used in writing and rewriting his dilemmas. Note that a critical feature of dilemma writing is that the dilemma needs to pit two conflicting ideals against each other (so that the solution is not merely finding an ideal, but involves resolving ideals that conflict).

It is often assumed that current controversial topics make the best dilemmas—the hotter and more talked-about, the better. Accordingly, old dilemmas—like the controversies about the Vietnam war—are worn-out dilemmas, distant from the subject's immediate experience, and hence are not good. To be sure, getting the subject's interest is important and interest is more likely with current, hot topics. However, the down side of using current, hot topics is that they are much talked-about, making it difficult to disentangle the subject's own structuring tendencies from verbalization coined by someone else. Remember, the assessment of moral judgment attempts to depict the natural ways that subjects make meaning of social events—their general default schemas—not a person's memory for verbalisms urged upon them by today's editorial, yesterday's TV show, or the required line of patter for the midterm in an ethics course. When topics are so current that slogans appear on bumper stickers, we are apt to get bumper sticker verbalizations for moral judgments.

Westbrook stated, "The challenge for the researcher in Journalism Ethics then, is differentiating between what is principled thinking concerning tough ethical dilemmas, and what are rotely acquired distillates, or worse, little-understood by-words." He talked about *formulaic shorthand phrases* that emerge from lengthy, complex legal decisions; about shorthand phrases becoming *distillates* of more extensive discussions; about bywords and catch phrases becoming learned by rote as *craft elements*, used to direct decisions reflexively. A major problem, then, in writing items or in interpreting interview material is that people may not mean the same thing with the same words.

There are several features designed into DIT items to minimize this problem:

1. Using short fragments as items rather than long speeches.
2. Using questions as items to raise issues instead of declarative propositions (in the hope that an issue raised as a question would be less apparent to those who don't have that issue in mind already).
3. Trying to equalize the items in length and complexity of syntax.
4. Introducing meaningless items (that have complex sentence syntax and specialized vocabulary) as an internal reliability check on the subject's test-taking set (for fuller discussion, see Rest, 1979, chap. 4).

Perhaps these devices will be useful in the new construction of instruments.



*Decision 2* involves setting the criteria by which the researcher determines that some forms of moral judgment are better (or more developmentally advanced, more philosophically defensible) than others. The DIT uses Kohlberg's theory of six stages essentially (although there are some differences, see Rest, 1979, chap. 2). Westbrook's Journalism Instrument also drew on Kohlberg's descriptors and Kohlberg's rationale for claiming a certain developmental order. However, Chang's TTMR did not use Kohlberg Stages as criteria for describing forms of moral reasoning, nor did she use Kohlberg's stage theory to argue that some forms of moral reasoning are better than others. Rather, she used discussions with other professionals, thus arriving at "expert judgment" to classify responses as high, medium, or low in development. Chang's approach to setting criteria, in effect, focuses on a more intermediate level of conceptualization than the more general Kohlbergian analysis (as discussed in chap. 1). It is similar to the approach taken by Bebeau in setting criteria, in that it uses the consensus of expert judgment rather than Kohlbergian stages.

From a reading about what goes on in current applied ethics courses and from looking at their required reading selections, it seems that many ethics courses focus on the intermediate level of conceptualization. In these courses, the concepts used to drive professional decision making are intermediate-level concepts (e.g., notions such as *informed consent*, *paternalistic deception*, and *privileged confidentiality*). Therefore, we need to develop tests of moral judgment that key on these intermediate concepts. Using the approaches by Bebeau and Chang for identifying consensual expert judgment for setting criteria, it seems that we have a lot of work to do in developing profession-specific, intermediate-level tests of moral judgment. The topics and evaluation criteria of those who currently teach ethics courses can be a place to start collecting dilemmas and criteria. Ethics instructors can provide the key intermediate-level concepts for the development of standard assessment instruments.

*Decision 3* involves having a strategy for building a case for the validity of a test. Building a case for validity means thinking about what studies can be done and what the findings have to show. In chapter 1, seven types of studies bearing on the validity of the DIT are listed. In effect, these can be viewed as validity criteria for the DIT. Other sets of validity criteria are, of course, possible. For instance, Chang (chap. 4) described the seven criteria she used for producing validity data for the Test of Teachers' Moral Reasoning (TTMR). The fundamental point here is that instrument development entails having a strategy to validate it and actually doing the studies.

Note that having criteria for validity enables us to know when our procedure is working. Otherwise, we devise instruments and have little basis for knowing if the instrument is really measuring what it purports to measure, or for knowing which part of the instrument is working better than another part. For instance, an instrument might contain two kinds of items: *High Items*, representing more defensible moral thinking; and *Low Items*, representing less defensible moral thinking. Then, if two groups of subjects are tested, one group being an *Expert* group, and the other, *Less Expert*, then there is a basis for determining if the instrument really generates

different scores for the two groups. In addition, there is a basis for examining specific items of the instrument. The items that are not working need redesigning. Thus, by this bootstrap method, cycles of instrument creation and empirical testing can produce a useful instrument.

*Decision 4* concerns the way a researcher combines information from many specific responses into an overall score or index. This procedure is *indexing*. For instance, Chang (chap. 4) in the TTMR produced four rankings from five stories for a total of 20 numbers. How she puts these 20 numbers together as a single, overall score that represents the subject's thinking is the issue of indexing.

A large part of Chang's chapter described experimenting with various ways of indexing. She doesn't just assume that the first idea that comes to mind for combining item responses is going to be the best index. Rather, she systematically uses her validity criteria to inform her which method of indexing consistently produces the most valid scores. She uses empirical research to shape instrument development.

Currently, Evens (Evens, in preparation) is conducting a dissertation that is looking into multidimensional scaling as a way of arriving at a better index for the DIT. A colleague, Mark Davison (Davison, 1977; Davison & Robbins, 1978), applied this methodology to DIT data. At that time, the database with which Davison had to work was limited, and, disappointingly, the index that came from Davison's work in the 1970s did not produce consistent improvement on the *P* score. Now, Evens is drawing upon the extensive database of over 58,000 DITs to retry this method. Her work is in progress, and it will take a little time before we know how this new index is working with the DIT. Nevertheless, other researchers may want to apply this scaling technique to their new measures (like the TTMR), especially if they use empirical data to determine the relative advancement of items.

Lastly—regarding new tests—recall that the Four Component Model indicated that other measures besides tests of moral judgment are necessary and relevant to a full assessment of moral development in the professions. Bebeau's work (chap. 7) on Component I: Moral Sensitivity, in the context of dentistry is the most sustained work on another component. Her chapter sketched out ideas for assessing other components as well. McNeel (chap. 2) also reported developing a measure of moral sensitivity for college students. Ideally, we would want to have measures of all four components customized for each profession—a big order for research in professional ethics.

## *LINKS TO BEHAVIOR*

---

What good is all this assessment of all these components if it does not predict to behavior? What good are courses in moral education that bring about gains in these instruments if the gains are not related to real life?

Predicting real-life behavior has long been a concern of research. Chapters 1 and 12 refer to hundreds of studies and reviews of these studies. The short

conclusion is that there is a consistent, statistically significant link of moral judgment with behavior, but the link is weak.

Chapters in this volume report significant links as well: Chang reports links of moral judgment with various aspects of school teaching; Bredemeier and Shields (chap. 10) report links with athletes' behavior; Self and Baldwin (chap. 8) report links with the behavior of medical doctors; Ponemon and Gabhart (chap. 6) report on the behavior of accountants; Sprinthall (chap. 5) reports on the behavior of high school studies; and Bebeau (chap. 7) reports on dentists who get into trouble with the ethics board.

Perhaps the most spectacular finding so far linking moral judgment to behavior is reported by Duckett and Ryden (chap. 3): First-year DIT scores of entering nursing students predicted an impressive correlation of .58 to clinical performance ratings of nurses in their later years. (This means that those students who had higher *P* scores in the first year were the students rated higher by their supervisors in their performance as nurses in a clinical setting.) A correlation of .58 is quite strong in social science research, and is all the more impressive considering that other variables do not predict this well.

In a different approach, Thoma's work (chap. 12) suggests that all these links might be strengthened by using the *U* score. The *U* score is derived from the DIT, requiring no additional subject time, but is computed from existing DIT data. Thoma reanalyzes five studies that previously linked the DIT's *P* score with behavior. He adds his *U* score to the *P* score, and substantially increases the link between the DIT (using both *P* and *U*) with behavior (see Fig. 12.1 showing increases in all five studies). In an appendix, Thoma specifies how to use the *U* score. There is no reason not to use Thoma's *U* score in all research linking the DIT with attitudes or behavior. It is quite likely that clearer, stronger results would result.

Ponemon and Gabhart (chap. 6) discuss a different approach to studying behavior:

1. Specify as the outcome variable one specific instance of professional decision making (e.g., fraud detection in reading financial documents, or under-reporting of time actually spent on a project).
2. Write different pieces of information that might be used in decision making. Give different pieces of information to different treatment groups.
3. Measure different subject characteristics (e.g., extent of professional experience, DIT score).
4. Determine the interaction of information with subject characteristics in producing the outcome professional decision.

This kind of study simulates the microprocess of professional decision making. Ponemon and Gabhart describe studies finding a significant interaction of *P* score with information variables in producing the outcome decision. This type of study should be used much more in other professional fields.

Finally, we call attention to work that attempts to study all components of the Four Component Model simultaneously, with the hope of increasing prediction to

behavior. Thoma discusses this approach. Bredemeier and Shields (chap. 10) propose organizing a program of research looking at four processes in three contexts for a matrix of 12 components.

## CONCLUSION

---

These chapters have suggested many possibilities for a program of research in professional ethics. The authors hope that researchers will be inspired to check out some of the references provided and begin the studies that will more thoroughly investigate the ethical development of professionals.

## REFERENCES

---

- Davison, M. L. (1977). On a unidimensional, metric unfolding model for attitudinal and developmental data. *Psychometrika*, 42, 523-548.
- Davison, M. L., & Robbins, S. (1978). The reliability and validity of objective indices of moral development. *Applied Psychological Measurement*, 2(3), 391-403.
- D'Souza, D. (1991). *Illiberal education: The politics of race and sex on campus*. New York: The Free Press.
- Duckett, L., Rowan-Boyer, M., Ryden, Crisham, P., M. B., Savik, K., & Rest, J. R. (1992). Challenging misperceptions about nurses' moral reasoning. *Nursing Research*, 41(6), 324-331.
- Evens, J. (in preparation). *Indexing moral judgment using multidimensional scaling techniques*. Unpublished doctoral dissertation, University of Minnesota, Minneapolis.
- Nokes, K. M. (1989). Rethinking moral reasoning theory. *Image: Journal of Nursing Scholarship*, 21, 172-175.
- Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students*. San Francisco: Jossey-Bass.
- Penn, W. Y., Jr. (1990). Teaching ethics—A direct approach. *Journal of Moral Education*, 19(2), 124-138.
- Penn, W. Y., Jr. (1992a). *A logic primer: Skills for critical reasoning*. Unpublished manuscript, St. Edward's University, Austin, TX.
- Penn, W. Y., Jr. (1992b). *Seeds of justice: A study of principled moral reasoning*. Unpublished manuscript, St. Edward's University, Austin, TX.
- Rest, J. R. (1979). *Development in judging moral issues*. Minneapolis: University of Minnesota Press.
- Rest, J. R. (1986). *Moral development: Advances in research and theory*. New York: Praeger.