This is to certify that the

dissertation entitled

EVALUATING MEDICAL ETHICS TEACHING

presented by

Kenneth Ross Howe

has been accepted towards fulfillment
of the requirements for

___Ph.D.___ degree in _Philosophy_ and
Teacher Education

Major professor

Robert E. Floden

Date_April 25, 1985___

0-12771

EVALUATING MEDICAL ETHICS TEACHING

By

Kenneth Ross Howe

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Departments of Philosophy and Teacher Education

352/758

To Paul

--whose soccer matches were a welcomed diversion

ACKNOWLEDGEMENTS

FOREWORD

The major problem surrounding the evaluation of medical
ethics teaching has been too little interaction between
ethics instructors and educational evaluators in working out
appropriate strategies for evaluatiing instruction; they go
their separate ways, each having little idea about the
nature of the others' activity. To remedy this problem,
this dissertation is aimed at both groups. Although this a
reasonable goal, anyone who has ever been involved in
interdisciplinary activities knows they create problems of
their own. In this dissertation, the attempt to speak to
both audiences periodically engenders the difficulty that
the same arguments which may appear too simplistic to one
group may appear too technical to the other. Though by no
means a total solution to the problem, I have tried to
lessen it by a liberal use of notes to amplify some points
and familiarize readers with others.

The arguments and findings of this dissertation have
broader applicablity than its title suggests. It is couched
in terms of medical ethics teaching rather than applied
ethics teaching in general for fortuitous reasons. Medical
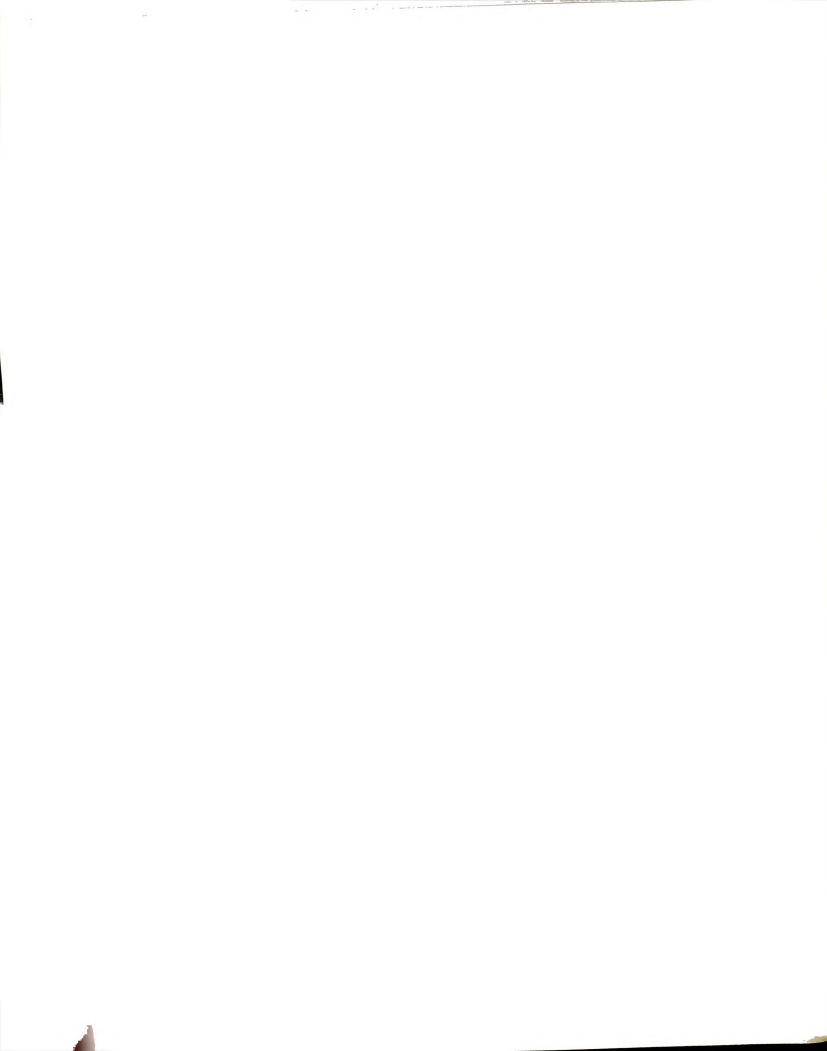and nursing ethics are where most of the activity has been

and are the areas in which I have done empirical research. Medical and nursing ethics, however, are just two instances of the more general field of applied and professional ethics. This connection is made explicit in the concluding chapter.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The Setting

Moral problems have been part and parcel of health care since the time of Hippocrates. But due in large measure to the emergence of specialized hospital care and strides in medical knowledge and technology, the frequency and complexity of such problems has steadily increased over the past several decades (Hunt and Aras, 1977). The general phenomena of growing pluralism, litigiousness, and consumerism have no doubt also been instrumental (Toulmin, 1981; Starr, 1982).

Although one might join McIntyre (1980) in lamenting the need to "rediscover" ethics, a burgeoning interest has occurred over the last decade and a half and has resulted in a significant rise in the number of individuals and institutions formally engaged in teaching medical ethics[1] (Clouser, 1980). Related to this, the American Association of Medical Colleges has recently moved to alter the course medical education has taken since Flexner, calling for major shifts of emphases in medical school curricula (1984). One of the three broad areas of curricula it recognizes, "Personal Qualities, Values, and Attitudes", is related to

1

ethics in an intimate way.

In light of these developments, it is unfortunate that no very satisfactory ways of evaluating medical ethics teaching have been devised. Few researchers publish empirical studies or develop instruments relevant to the evaluation of medical ethics teaching[2]. This paucity of research reinforces one source of resistance by administrators and students to medical ethics, namely, doubt about whether there are appropriate means for evaluating ethics instruction (Hastings Center, 1980). Skepticism about evaluation is in turn rooted in a set of beliefs about ethics itself. As enumerated by Callahan (1978), these beliefs include the notions that ethics is soft, subjective, unscientific, indoctrinating, pedantic hair-splitting, irrelevant to the real world, and unteachable.

Resistance by ethics instructors to evaluation, it seems, can only strengthen the very views they spend much of their time trying to undermine. Evaluation of innovative educational programs has become a fact of life.[3] If proponents of medical ethics teaching wish to hold their present ground and to make further inroads in already crowded medical and nursing school curricula, then they need to develop means of assessing their curricula and demonstrating its value. Less prudentially, well-conceived evaluation can serve to enhance the quality of instruction. Because evaluation is essentially systematic and informed criticism, properly done, it can help locate successful and

unsuccessful modes of instruction and prompt needed
improvements.

The Task

The general aim of this dissertation is to develop a
defensible strategy for evaluating medical ethics teaching.
This will require satisfying demands and meeting criticisms
from two quite different perspectives. On the one hand, the
strategy must incorporate methods and criteria which ethics
instructors can endorse. On the other hand, the methods and
criteria must be anchored in a defensible position on
applied social research. To date, meeting one set of
demands has precluded meeting the other, rendering eval-
uation of medical ethics teaching trivial, impressionistic,
or both.

Evaluation is a complex and controversial task under the
best of circumstances. Evaluating medical ethics teaching
is especially problematic because evaluators and medical
ethics teachers clash over the proper means (or over the
very possibility) of evaluating ethics teaching.

Those engaged in medical ethics teaching have found
evaluation strategies ill-conceived for a variety of
reasons. Criticisms have been advanced both from those
whose projects have been evaluated (Ruddick, 1981) and from
others attempting to forestall the use of questionable
criteria (e.g., Clouser, 1973 and 1980; Callahan, 1980; and
Goodpastor, 1982). These critics share the belief that
evaluation is often far removed from the specific concerns

and goals of medical ethics teaching. They have urged that
attempts to produce changes on measures borrowed from moral
psychology can actually undermine ethics teaching and that
attempts to produce changes in moral behavior can themselves
be morally objectionable.

According to Caplan (1980), resistance to evaluation is
pervasive among ethics instructors, and such resistance is
at least partially justified given evaluation's track
record. But part of the problem may be attributed to ethics
instructors' sometimes strident rejection of the methods of
evaluation. For example, Clouser quips, "An evaluation too
simple and mundane to be of interest to the statistician,
but of tremendous help to the teacher of the course, is the
evaluation form filled out by the student at the conclusion
of the course" (1980, p. 32). Clouser's remark is naive;
student evaluations are taken seriously by educational
evaluators ("statisticians") and a fair amount of research
has been conducted in this area[4].

On the other hand, part of the problem may be attributed
to evaluators. Many of them employ a pair of Procrustean
distinctions--between facts and values and between
qualitative and quantitative methods--that are obstructive
when evaluating ethics teaching. Facts and quantitative
methods are associated with objectivity and science; values
and qualitative methods, with subjectivity and non-science.
Ethics falls squarely within the subjective, non-scientific,
qualitative domain of values. By implication, ethics

instruction is evaluable only in a "soft" sense in which cataloging changes in values is the major focus. Ethics instructors rightly reject such a naive construal of the fact-value distinction and the evaluation approaches which grow out of it.

Although there is some evidence for the belief that evaluation research is hopelessly muddled and therefore bound to distort and trivialize the aims of ethics teaching, the conclusion that more rigorous and systematic evaluation is impossible is hasty at best. The shortcomings of past evaluations result from inappropriate methods, criteria, and instruments (criticisms by no means confined to ethics instruction as an object of evaluation), not from the attempt to be rigorous and systematic per se. While the precise nature of ethics is an important consideration in developing a general evaluation approach, there is nothing inherent in ethics instruction that precludes systematic evaluation.

Structure of The Dissertation

The basic framework of the dissertation is as follows: (1) A set of goals for medical ethics teaching is developed to serve as the basis for evaluation. (2) A general methodological approach that emphasizes generating useful information (instead of emphasizing methodological rigor for its own sake) is defended against positivist-inspired criticisms. (3) A model for the evaluation of medical ethics courses is developed on the basis of the goals and

the general methodological approach. (4) Three concrete
examples of course evaluations are used to illustrate and
provide a test of the model. (5) The model is further
examined and suggestions are offered for extending it to
medical education formats besides traditional courses
and for extending it to applied and professional ethics more
generally.

Goals for medical ethics teaching are addressed in
Chapter II. A set of seven goals is developed largely on
the basis of the John Wilson's "components of the morally
educated" (1967). The goals are distinguished into "direct"
and "indirect" varieties, reflecting the respective moral
and practical differences between educational aims such as
imparting knowledge on the one hand and promoting courage on
the other. The goals are then evaluated in terms of their
ability to withstand criticism and in terms of the degree to
which they are representative of the goals of ethics
teaching which experts establish for themselves. Chapter II
emphasizes the importance of clarifying what is to be
evaluated in medical ethics teaching.

Chapter III addresses general issues of methodology
associated with how to evaluate medical ethics teaching.
Because such fundamental issues are involved, the arguments
in Chapter III at times lead far beyond the specific case of
medical ethics teaching evaluation. In particular, ethics
has, since the advent of positivism, been relegated to the
"soft" side of the hard-soft dichotomy of knowledge.

Because the subject itself is viewed as soft, many educational researchers believe means of evaluating it must also be soft.

The general aim of Chapter III is to criticize the hard-soft dichotomy and the notion that social research might or should be value-free. Sound criticism of these related notions is required to rescue both ethics and appropriate evaluation methods from unwarranted and misguided attacks. The fact-value and quantitative-qualitative distinctions are considered in some detail. The advisability of appealing to the methods and concepts of moral psychology is also considered. The chapter shows that the drive to be rigorous and scientific (i.e., to disparage so-called soft knowledge) is based on untenable positivistic strictures and that following these strictures compromises evaluation's value. Achieving worthwhile purposes, rather than obeisance to a priori standards of methodological rigor, is the legitimate criterion for conducting and assessing evaluation.

These general methodological considerations pave the way for Chapter IV, which characterizes a "basic functional model" for the evaluation of medical ethics courses and then illustrates how the model may be used to interpret three course evaluations. The examples illustrate how the more general and more-or-less conceptual considerations of the preceding chapters embodied in the basic functional model can be used to inform evaluation practice. The impact of

the findings of the three evaluations on teaching practice
is also discussed, and is related to the credibility of the
evaluation model employed.

The concluding chapter uses the three examples of
Chapter IV to support general theoretical claims about the
goals advanced in Chapter II and the methodology advanced in
Chapter III. The concluding chapter also suggests how the
general approach defended and illustrated in the previous
chapters might be extended to medical ethics teaching beyond
the limited context of courses considered in Chapter IV.
Problems that can be anticipated and changes needed in the
nature of medical education are broached. How the
conclusions of the dissertation might be extended to other
areas in applied and professional ethics is also discussed.
Finally, three explicit disclaimers about the scope of the
dissertation are made.

Development of the Arguments

The chronology of thought is not the neat one suggested
by the outline above. The arguments about methodology in
Chapter III were at best inchoate, and those about goals in
Chapter II were non-existent when the first concrete example
of Chapter IV was carried out in 1980. By the time the
second course evaluation was accomplished in 1982 progress
had been made on these fronts but it was slight. The
thought reflected in Chapters II and III did not guide the
evaluations until the study of 1983.

The development of thought was thus back and forth

between practice and more general theoretical concerns. This has certain advantages. It provides a way of testing theoretical views against their practicability. This in turn leads to revisions in the general stance and a better approach on subsequent occasions. Implementation in practical settings also provides a means to detect interests, misunderstandings, and sources of resistance. This, too, informs the shape of the general theoretical stance. On the other hand, this chronology of thought has a certain disadvantage related to the general structure of this dissertation. In so far as the concrete examples illustrate and provide a test of adequacy of the general approach, at times they have to be shoehorned into the conceptual framework established. This problem is not insuperable, but is worthy of note. It should be kept in mind when reading Chapter IV, which by and large is a post hoc assessment of concrete evaluations in terms of a conceptual framework that did not dictate the ways in which they were designed and conducted.

NOTES

1. Unless the more specific meaning is made clear by the context, 'medical ethics' will be taken throughout to include nursing ethics.

2. A 1982 MEDLINE search failed to produce any useful instruments or methods. Stolman et al. (1982) and Siegler et al. (1982) report similar difficulties.

3. Passage of the Elementary and Secondary Education Act (ESEA) in 1965 is the generally agreed upon benchmark of the Federal Government's entry into public education (e.g., Worthen and Sanders, 1973). With ESEA came significant federal funding and the requirement for evaluation.

4. For a review of the research, see Aleamoni (1983).

CHAPTER II

EXPLICATING THE GOALS OF MEDICAL ETHICS TEACHING

LAYING THE FOUNDATION FOR CREDIBLE EVALUATION

> We have to start with conceptual questions, then move on to the
> empirical theories...then on to educational practice in the
> classroom. If we do not get the concepts and categories clear
> in the first place, we shall not know what or what sorts of
> facts, theories and practice we ought to look at. (Wilson,
> 1983, p.192)

This chapter aims to develop a set of goals to frame the

evaluation of medical ethics teaching. As Wilson suggests,

establishing a conceptual framework is the necessary first

step: it provides the concepts and categories needed to

organize and direct data collection and interpretation. In

so far as it is the first step, the credibility of a

framework of goals for medical ethics teaching is a crucial

determinant of the credibility of evaluation.

The chapter is divided into two major sections that

correspond to two ways in which the credibility of the

framework will be established. First, goals need to be

credible to ethics teachers. A set of seven goals will be

specified and then compared to goals endorsed by leaders in

the field of medical ethics. In light of the resistance of

ethics instructors to evaluation discussed in the first

chapter, it is important to acknowledge the goals medical

11

ethics teachers adopt for themselves. Otherwise, medical
ethics teachers will continue to dismiss evaluation as
irrelevant or destructive. Second, goals need to be
credible to a broader audience that includes educational
researchers and health care professionals engaged in
education. Critical discussions of "controversies" and
"misconceptions" will be provided to help blunt the
objection that evaluation in terms of the avowed goals of
medical ethics teachers embodies a merely conventional,
unexamined, and partisan perspective and set of interests.

Specifying the Goals

Developing a set of defensible and practicable
educational goals requires adjusting means and ends--the end
of roast pig, for example, looks silly if the means employed
are burning down the barn (Dewey, 1939). The specification
of the goals of medical ethics teaching will be accomplished
in four steps. (1) Competence in medical ethics, the ideal
end, will be described. (2) This ideal end will then be
considered in terms of moral and practical constraints
imposed by the educational context; "direct" and "indirect"
goals will be distinguished as a way of responding to these
constraints. (3) Seven goals will be advanced. (4) These
seven will be compared to two current, representative views
of the goals of medical ethics teaching.

John Wilson's work (1967, 1969, and 1973) provides the
point of departure. A philosopher, teacher, and prolific

writer in the fields of moral philosophy and educational research methodology, Wilson spent 10 years directing the Farmington Trust project in moral education. The project included empirical research as a major facet, and Wilson had primary responsibility for developing philosophically informed criteria and research methods to guide the research. Although his concerns were more general and detailed than the ones here, his "components of the morally educated" are readily adaptable to medical ethics teaching.

Competence in Medical Ethics: Wilson's
Components of the "Morally Educated"

Wilson develops a set of traits that he labels the "components of the morally educated". Each of the components captures a dimension of the "morally educated" which is (1) necessary, and (2) logically independent of the other components. These components (capitalized throughout to indicate the intended usages) are the following: Moral Regard, Empathy, Interpersonal Skills, Knowledge, Reasoning, and Courage.[1] The case described below will be used to illustrate Wilson's components.

> Margaret Scanson, North Lake Community clinical nursing instructor, is presently supervising students at Portage City Memorial Hospital. Common practice in the hospital is for nurses not to tell patients whether or not they have cancer, because some doctors prefer that they not know. Margaret has suggested to her students that if a patient asks such a question, the student should ask the patient what the doctor has said, and if the patient wishes, should offer to speak with the doctor about the matter.

Margaret assigned a student, Marie Blanchard, to care for Mrs. Bullough, a woman in her early thirties who had a brain tumor. Marie, in the course of being with Mrs. Bullough prior to surgery, observing the surgery, and caring for her afterwards, learned that the tumor was malignant. When Marie arrived the following day to care for her, Mrs. Bullough, who knew Marie had been with her in surgery and the recovery room, immediately asked, "Is it cancer?" Marie was at a loss because the stock answer, "What did your doctor say?" seemed such a denial of Mrs. Bullough's need for an answer, and since Mrs. Bullough had good reason to believe that Marie knew the answer, she couldn't blandly say she didn't know. Marie excused herself from Mrs. Bullough and sought her instructor.

After learning that the doctor normally shared a diagnosis of malignancy with his patients but that he would not be available until later in the day, Marie questioned the value of ignoring Mrs. Bullough's concerns. She requested that Margaret allow her to tell the patient or, if that was unacceptable due to her inexperience, that the head nurse or Margaret herself tell Mrs. Bullough that the tumor was malignant. Both Margaret and the head nurse knew the diagnosis and the planned treatment. In addition, both nurses had spent time with Mrs. Bullough during the diagnostic period prior to surgery and probably knew her as well or even better than the surgeon. What should Margaret do?

According to Wilson, Margaret would have to manifest the six components mentioned above to work her way through this problem. Moral Regard is the most fundamental and involves counting the interests and feelings of others as equal to one's own. Margaret would have to recognize that there are individuals whose interests might conflict--hers, the student's, the physician's and Mrs. Bullough's, to name the most salient ones--and that each of these individual's interests carries the same initial weight. There simply would be no moral problem for Margaret otherwise.

Margaret would also have to be empathetic. She would have to be able to recognize other's feelings and be able to correctly describe them. For example, is Mrs. Bullough

merely frightened, or is she so distraught that her
competence might be questioned? Margaret could go wrong,
for instance, if she described Mrs. Bullough as hysterical
when she was instead understandably disconcerted.

The third component that Margaret would have to exhibit
is Interpersonal Skills. She would have to be able to
correctly interpret Mrs. Bullough's facial expressions, tone
of voice, and posture. She would have to ascertain Mrs.
Bullough's values, sincerity, her understanding of the
situation, and the appropriate language to use for effective
communication. The manner in which she responded to Mrs.
Bullough--that is, Margaret's own facial expressions, tone
of voice, posture--would of course also have important
consequences for the quality of the interaction. And the
importance of Margaret's Interpersonal Skills would not be
limited to her exchanges with Mrs. Bullough; it would also
determine her effectiveness in dealing with other members of
the health care team whose interests and feelings also would
be relevant.

Next, Margaret would have to possess Knowledge which
would allow her to formulate reasonable strategies and to
anticipate the consequences of her actions. Would telling
Mrs. Bullough do more harm than good? Is there evidence
supporting the claim that cancer patients would, despite
their avowals, really prefer not to be told? Is there some
especially effective way to tell them? What if, when told,
Mrs. Bullough immediately requests information regarding her

treatment and prognosis that outstrips Margaret's expertise?
And so on.

Margaret would also have to be able to draw conclusions
on the basis of the preceding components in order to derive
some rule of conduct to apply in this case or to recognize
the situation as an instance of one of her previously
derived moral rules. In short, she would have to do some
reasoning. For example, suppose she endorsed the following
rule: Health care professionals are obligated to disclose
information to patients that has significant impact on their
life plans unless there are compelling reasons for not doing
so. Suppose in addition that she is convinced that Mrs.
Bullough is rational and sincere in her request. Other
things equal, Margaret should be able to combine her
assessment of Mrs. Bullough and her principle and draw the
conclusion that Mrs. Bullough should be told.

The last component Margaret would have to exhibit is
Courage. Believing that Mrs. Bullough should be told is one
thing, actually taking steps to ensure this occurs is quite
another. Pursuing the matter at all entails certain risks.
Rocking the boat could result in alienating her co-workers,
moral criticism, or even the loss of her job. If these
hurdles are cleared, telling Mrs. Bullough entails an
emotionally trying, perhaps painful, experience.

With Margaret's dilemma now in hand, Wilson's schema of
components may be summarized as follows: Moral Regard is
the fundamental presupposition of moral behavior and

decision-making. Empathy and Interpersonal Skills are required to ferret out and articulate the interests and feelings that others have. The results of the operation of these components informs beliefs about given individuals and is combined with more general background knowledge and beliefs. Information from all these sources is then used to reason through to a principle of action. Finally, Courage is required to convert conclusions to actions where difficult situations are involved.

Accommodating Moral and Practical Constraints: Distinguishing Direct and Indirect Goals

Wilson's components are constitutive of "morally educated" behavior, but they may not thereby be straight-forwardly converted into educational goals. In order to develop goals based on Wilson's components that are consistent with constraints on formal education, it will be helpful to construe Knowledge and Reasoning as cognitive characteristics necessary for knowing how one ought to act and to construe Moral Regard, Empathy, Interpersonal Skills, and Courage as personality characteristics necessary for acting in conformity with knowing how one ought to act (e.g., Frankena, 1975). This two-way division of the components, though crude, will help in the task of specifying appropriate educational goals.

In an explicitly educational context (versus a therapeutic one, for instance), goals associated with

cognitive characteristics differ in both practical and moral dimensions from goals associated with personality characteristics. The practical difference is that personality characteristics are affected in profound ways by exogenous and non-rational causes which are unknown to teachers and students and largely beyond their control-- rearing, socio-economic status, health, and genetic make-up are a few examples. Personality characteristics are therefore significantly resistant to change, especially via brief educational experiences. **All things equal**, cognitive characteristics are amenable to change through standard educational techniques (e.g., lectures and discussions) and are under much greater control of educators.

The moral difference between cognitive and personality characteristics involves the relationship between means and ends in education. Put crudely, education should aim at the rationality of the student; it should consist in giving reasons in conjunction with rational arguments, not in things such as coercion and manipulation (e.g., Scheffler, 1978; Peters, 1967; and Wilson, 1967). Again, **all things equal**, ends associated with cognitive characteristics are well-suited to means such as lectures and discussions that aim at the rationality of students. By contrast, ends associated with personality characteristics are less subject to influence through these rational means. Such ends are often more closely associated with techniques

that are potentially objectionable within the educational context, such as behavior modification and the "therapeutic manipulation" which Michael Scriven (1975) associates with "affective" moral education.

Some (and perhaps many) ethics instructors conclude that these moral and practical differences between cognitive and personality characteristics entail that the latter are beyond the scope of ethics teaching. Such an extreme position is unwarranted. Presumably, no one thinks it is impossible for a course in ethics to improve students' ability to empathize. Furthermore, if two courses were otherwise the same, the one that improved Empathy would be judged superior; and the same could be said of Moral Regard, Interpersonal Skills, and Courage. It is therefore a mistake to altogether eliminate these personality characteristics from among the goals of ethics teaching.

A less extreme way of accommodating the constraints that the educational context imposes is to distinguish between "direct" and "indirect" goals. Goals associated with relatively unproblematic cognitive characteristics-- Knowledge and Reasoning--may be considered "direct" goals. Goals associated with relatively problematic personality characteristics--Moral Regard, Empathy, Interpersonal Skills, and Courage--may considered "indirect" goals. The distinction between direct and indirect goals is designed to acknowledge the practical problem presented by personality characteristics (i.e., they are difficult to influence) and

thus signals the degree to which responsibility may be assigned for achieving goals associated with them. Medical ethics instruction is uniquely responsible for a direct goal such as imparting the Knowledge peculiar to its domain; it is responsible for promoting an indirect goal like Courage only partially and only as one influence among many. The distinction between direct and indirect goals also signals a genuine moral difference between the legitimate means by which student behavior may be influenced; it separates those goals most easily associated with education from those which potentially involve paternalism and manipulation. Acknowledging this difference in means is especially important in medical and nursing education where students have the full range of rights and expectations of autonomous adults.

The Wilson+1 Goals: Seven Goals for
Medical Ethics Teaching Described

Employing the direct-indirect goal distinction, this section advances seven goals for medical ethics teaching. "Appreciation", a goal not associated directly with Wilson's components, combined with "imparting Knowledge" and "improving Reasoning", form a set of three direct goals. "Stimulating Moral Regard", "eliciting Empathy", "enhancing Interpersonal Skills", and "promoting Courage" form a set of four indirect goals.

Direct Goals

Imparting Knowledge. Imparting Knowledge is a familiar and straightforward educational goal. In medical ethics, it consists in imparting the facts, concepts and positions that are peculiarly medical-ethical and especially pertinent to moral problems in medicine.

Improving Reasoning. In contrast to Knowledge, Reasoning is an exceedingly complex concept, and moral philosophers vigorously resist reducing Reasoning to static principles or rules that can be precisely specified ahead of time. For instance, Goodpastor (1982) criticizes the Kohlbergian psychologist James Rest (1982) for suggesting that ethics teaching might be evaluated in terms of student progress relative to pre-established "stages" of moral judgment. According to Goodpastor, pre-set criteria ignore moral reasoning's self-critical and evolving nature. Dewey makes a similar point when he describes philosophy (including moral philosophy) as "thinking which has become conscious of itself" (1944, p. 326). Finally, MacIntyre (1981) contends that the quality of moral argument must be judged in terms of evolving criteria "internal" to the "practice" itself, and not in terms of criteria which are static and "external".

Most moral philosophers would agree with Goodpastor, Dewey, and MacIntyre that pre-set, rigid standards of correct moral argumentation cannot exist. Despite the self-critical and evolving nature of moral reasoning,

however, moral philosophers (and ethics instructors) are able to distinguish good from bad arguments and to articulate general rules of thumb. One such set of rules (designed for medical ethics in particular) is suggested below. Although it could probably be improved, it provides the reader with the flavor of what is meant by Reasoning.

1. recognizing ethical problems and formulating them in terms of the relevant issues involved

2. engaging in conceptual analysis by
   a. drawing necessary and relevant distinctions
   b. clarifying important concepts which are vague or ambiguous

3. distinguishing the following dimensions of medical-ethical decision-making:
   a. legal
   b. medical-technical
   c. resource allocation

4. formulating arguments that are
   a. clear
   b. consistent
   c. logically correct
   d. factually correct

   and that

   e. identify alternative or competing positions
   g. identify presuppositions in various positions
   g. anticipate and address objections

Instilling Appreciation. Wilson's components may well implicitly include some notion of appreciation. It is sufficiently important, however, to be a distinct goal.

For the purpose of illustration, imagine medical students A, B, and C are enrolled in a medical ethics course and fit the following descriptions. Student A does extremely well on the written exams and exhibits good verbal facility regarding the key concepts and arguments, but A

judges medical ethics to be a waste of time and just another educational hoop to jump through. Student B is enthusiastic about the course, but, unlike A, has much difficulty with the exams and cognitive aspects of the course in general. Although B is favorably disposed, B is unable to clearly articulate what value medical ethics has. Student C blends A's cognitive skills with B's favorable attitude, and C claims that medical ethics teaching is valuable because it stimulates thinking and prompts an awareness of alternative views.

Given the intended sense of the term, student A does not Appreciate the course (for simplicity, Appreciation of the course is identified with Appreciation of ethical inquiry). Though successful with regard to cognitive skills, A denies the course has value. A is either unable to or refuses to see the point. B also does not Appreciate the course. Though possessing a positive attitude, B does not provide evidence of a clear enough understanding to infer that the positive attitude applies to the correct object. In this way B also fails to see the point. Only C can be said to Appreciate the course (i.e., to Appreciate ethical inquiry) in the intended sense of the term. Unlike B, C knows how to play the game according to its cognitive rules; unlike A, C values what the game is about.

Appreciation is important because ethics teaching that involves students merely in intellectualizing on the one hand or merely in emoting on the other is incomplete and

distorted. Students need to value the importance of rational inquiry in ethics in a way that rivals the value placed on cognitive investigation in other aspects of clinical decision-making. As Clouser puts it, it is essential to "convey that this enterprise [medical ethics] is not just a mind game, but deals with matters of profound significance to self, society, and profession" (1980, p. 18). Simply getting students to want, like, or enjoy medical ethics teaching--getting them to be like student B--is not sufficient. To Appreciate ethical inquiry, students must, like student C, know why ethical inquiry is valuable, which in turn requires that they be able to successfully engage in it. Appreciation, then, has both cognitive/skill and affective/attitudinal dimensions. The educational importance of Appreciation is that its presence increases the probability that students will act in accord with the intended lessons of medical ethics teaching.

Indirect Goals

Indirect goals are distinguished from direct goals in terms of the moral and practical constraints that apply more forcefully to the former. Because these constraints apply more forcefully to the indirect goals, medical ethics teaching should, all things equal, produce improvement in terms of the direct goals, but is required only to draw students out--to reinforce or stimulate their development-- in terms of the indirect goals. Viewed in this way, direct goals may be associated with final outcomes that

medical ethics teaching should achieve, and indirect goals may be associated with process outcomes that medical ethics teaching should prompt.

To illustrate how the indirect goals--Moral Regard, Empathy, Interpersonal Skills, and Courage--may be considered goals of the process of medical ethics teaching, the remainder of this section provides brief discussions of each of the indirect goals to show how they might be pursued "indirectly" (i.e., as part of the process of medical ethics teaching).

Before turning to these illustrations, the reader should note that the verbs used in the goal-statements themselves reflect the direct/indirect distinction. The verbs 'imparting', 'improving', and 'instilling' are used in connection with the direct goals. Borrowing somewhat from Callahan (1980), the weaker verbs 'stimulating', 'eliciting', 'reinforcing' and 'promoting' are used with the indirect goals.

Stimulating Moral Regard. Persons altogether lacking in Moral Regard (sociopaths, for instance) have no business in medical and nursing schools; no educators can be expected to make much progress with such individuals. Short of such extremes, however, medical ethics instruction should bear some responsibility for stimulating the Moral Regard of those who possess it and for stimulating them to be reflective about their moral views.

Consider the common educational method of discussing

ethically problematic medical cases. The mere undertaking of such discussions stimulates Moral Regard because ethical issues, by their very nature, begin with Moral Regard--i.e., without a perceived conflict of interests and the desire to adjudicate the conflict, no moral puzzlement exists. Because Moral Regard is so fundamental, then, simply engaging students in moral inquiry in general is bound to stimulate Moral Regard.

On the other hand, Moral Regard may be stimulated at more specific levels--and as a consequence of the process of pursuing the goals of Knowledge and Reasoning. For instance, clarifying for medical students the religious beliefs underlying the refusal of a blood transfusion by a Jehova's Witness stimulates students' Moral Regard for the individual refusing treatment and for persons with unconventional views more generally. The clarification of the patient's beliefs per se is cognitive and educational, and stimulating students' Moral Regard is an outcome of this educational process. Stimulating Moral Regard by these "indirect" educational means may be distinguished from the causal means that would be involved, for instance, in a "values clarification" exercise designed to directly (and surrepticiously) bring about greater regard for others.

Eliciting Empathy. The ability to assume the viewpoints and feelings of others is essential in providing the insight necessary to intelligently work through ethical problems. The close connection between reasoning and feelings

(feelings are a kind of "data") needs to be brought home to students, lest ethics teaching become incomplete and sterile. However, "touchy feely rap sessions", emotionally charged "encounter sessions", and other non-cognitive, causal means of affecting Empathy are out of place in the educational setting. Instead, conventional educational strategies such as role-playing, viewing video tapes, and simulated patient encounters should be employed to elicit Empathy. Simply discussing cases can serve the aim of eliciting Empathy, but probably less effectively.

Reinforcing Interpersonal Skills. Interpersonal skills are required in virtually all aspects of patient care. A physician or nurse who lacks such skills is handicapped, because the skills are required in the most rudimentary tasks, such as reducing anxiety, conveying information, or obtaining information. Although teaching Interpersonal Skills has become a legitimate direct goal of medical education (e.g., Kahn et al., 1979), it is unreasonable to require medical ethics teaching to adopt Interpersonal Skills as more than an indirect (or process) goal.

The special charge of medical ethics instruction with respect to Interpersonal Skills is reinforcing students' ability to interact with both their patients and colleagues in connection with often difficult and emotionally charged moral issues. Role-playing, video tapes, simulated patient encounters, and discussion again come to mind as educationally acceptable methods to achieve this goal.

<u>Promoting Courage</u>. As is true of the other indirect goals, medical ethics instruction cannot be expected to shoulder the full responsibility for altering a trait acquired over many years and resulting from causes that are poorly understood. Like Moral Regard, however, Courage (or the relationship between avowals and actions) is a traditional topic in ethics. In connection with medical ethics in particular, there is a strong tendency among students and professionals to look to conventional practice as the standard of appropriate behavior. The claim is sometimes made that arguments of ethicists are all fine and good but are not "realistic". Courage may be promoted by pointing out that being "realistic" may be just shorthand for being timid, or even egoistic, and that one is morally obligated to act in accordance with one's well-considered moral judgments. In a more general vein, working through ethical problems and developing rational justifications for ethical viewpoints--characteristics of <u>educational</u> methods--promote Courage by providing students with the confidence in their moral beliefs they need in order to convert such beliefs into actions.

Representativeness of the Wilson+1 Goals

Seven goals for medical ethics teaching have been developed based largely on Wilson's "components of the morally educated". This section compares these seven to goals endorsed by experts on medical ethics teaching.

Although the issue of goals for medical ethics teaching has generated little careful discussion, two proposals merit examination: Callahan's (1980) goals and the 1983 DeCamp conference goals.

Callahan's five goals for ethics teaching were prominent in the Hastings Center's investigation of ethics in higher education. Caplan employed them in the volume which resulted, Ethics in Higher Education (1980) when discussing the evaluation of ethics teaching in general. Clouser (1980), in another Hastings Center volume, employed them in his discussion of evaluating medical ethics teaching in particular. The prominence in the field of medical ethics of the Hastings Center and of these individuals indicates that Callahan's goals are reasonably representative.

Callahan's five goals are: stimulating the moral imagination, recognizing ethical issues, eliciting a sense of moral obligation, developing analytical skills, and tolerating--and reducing--disagreement and ambiguity. Callahan also includes "context-dependent" goals, which allow for the special content of medical ethics. The mapping in Figure 1 is a suggested interpretation of Callahan's goals in terms of Wilson's+1.

The more recent DeCamp conference involved eleven leaders in the field of medical ethics.[5] Four general goals resulted from their deliberations, and a fifth, paralleling Callahan's "context dependent" goals, specified the minimum of issues (i.e., content) which any medical ethics teaching

Callahan's Goals		Wilson+1 Goals
stimulating the moral imagination	---->	stimulating Moral Regard, eliciting Empathy
recognizing ethical issues	---->	stimulating Moral Regard, eliciting Empathy, imparting Knowledge
eliciting a sense of moral obligation	---->	promoting Courage
developing analytical skills	---->	improving Reasoning
tolerating—and reducing—ambiguity and disagreement	---->	enhancing Interpersonal Skills, improving Reasoning
context dependent goals	---->	imparting Knowledge

Figure 1. A Comparison of the Callahan and Wilson+1 Goals

program should address. The four general goals, A-D, are

A. clarification of central concepts (e.g., competence);

B. understanding of important decision-making procedures (e.g., when is it morally justified to treat an unwilling patient?);

C. ability to apply concepts and decision-making procedures to actual cases;

D. various interactional skills (e.g., discussing with a terminally ill patient her wishes about going on Do Not Resuscitate status).

Like Callahan's, these goals are consistent with Wilson's+1; Figure 2 compares the two sets.

DeCamp Group Goals		Wilson+1 Goals
clarification of central concepts	---->	imparting Knowledge, improving Reasoning
understanding of important decision-making procedures	---->	imparting Knowledge, improving Reasoning
ability to apply concepts and decision-making to actual cases	---->	all
interactional skills	---->	stimulating Moral Regard, eliciting Empathy, enhancing Interpersonal Skills, promoting Courage

Figure 2. A Comparison of the DeCamp Group and Wilson+1 Goals

Callahan's goals and those of the DeCamp conference exemplify the goals of medical ethics teaching endorsed by those presently engaged in and knowledgeable about medical ethics teaching. The consistency between these and Wilson's+1 exemplified in the mappings of Figures 1 and 2 indicates that Wilson's+1 coincide with current thinking in the field.

Advantages of The Wilson+1 Goals

Having established the representativeness of the Wilson+1 goals, the four-step plan for specifying the goals of medical ethics teaching is complete. Before turning to a critical examination of the Wilson+1 goals, three reasons for preferring these goals over the DeCamp

and Callahan alternatives are given.

Wilson developed his "components of the morally educated" with an eye toward empirical research, and the Wilson+1 goals have been developed with the same thing in mind. The aims of Callahan and the DeCamp group, by contrast, were to reach a consensus on goals that could be used to guide teaching and that could be effectively communicated to others. It should not be surprising that the Wilson+1 goals have several advantages over the DeCamp and Callahan goals for the purpose of framing the evaluation of medical ethics teaching. (1) The Wilson+1 goals are more discrete; this characteristic, reflected in the comparisons in Figures 1 and 2, facilitates focusing empirical research and aids the development of research methods and instruments. (2) The Wilson+1 goals explicitly distinguish direct and indirect goals. This helps establish priorities and reasonable expectations for medical ethics teaching. (3) The Wilson+1 goals explicitly include Appreciation. Appreciation is important in its own right and is readily amenable to investigation by the use of student evaluations of teaching.

Controversies and Misconceptions

As stated at the beginning of this chapter, the credibility of a given framework of goals for medical ethics teaching depends on (1) the degree to which the framework is representative of the views and practices of those engaged

in medical ethics teaching, and (2) the degree to which the framework withstands critical scrutiny. Representativeness of the Wilson+1 goals was established in the preceding section. This section critically examines the Wilson+1 goals by entertaining "controversies" and "misconceptions".

As used here, "controversies" involve disagreement about the nature of medical ethics teaching and the direction it should take. Four such controversies will be discussed: moral behavior as a goal, Appreciation as a direct goal, medical ethics as "formal", and ethical theory's place in medical ethics teaching. In contrast to the controversies, which take for granted the legitimacy of medical ethics teaching and disagree about the best approach, "misconceptions" involve the contention that medical ethics teaching is unnecessary or objectionable. Misconceptions are associated with popular and relatively unsophisticated views of medical ethics teaching, views that may be associated with groups such as students, health care professions faculty, and evaluators. Seven such misconceptions will be considered: ethics is subsumed by the social sciences; ethics is simply interpersonal skills; professional ethical codes are a sufficient means of dealing with ethical problems; legal considerations preclude ethical ones; ethics teaching conflicts with religious beliefs; ethics teaching is indoctrination; and formal education has no proper role to play in moral education.

Controversies

Moral Behavior as a Goal

The Wilson+1 goals (except for Appreciation) result from
analyzing moral behavior into its "components", and thus the
ultimate goal associated with the Wilson+1 framework is
moral behavior. On the other hand, philosophers often
eschew moral behavior as a goal of medical ethics teaching.
Callahan (1980), for instance, claims that moral behavior is
a "dubious" goal of medical ethics teaching, and his view is
echoed by others (e.g., Macklin, 1980, Caplan, 1980, and
Goodpastor, 1982). This apparent disagreement between the
view Callahan represents and the Wilson+1 goals can be
removed by observing that rejecting moral behavior as a goal
of medical ethics teaching is justified provided it is
identified with the behavioristic sense of overt behavior,
or provided that it is construed as the proximate goal of
medical ethics teaching which ethics teachers are to be held
directly accountable for.

Following a brief discussion of why philosophers are
correct to reject moral behavior in its behavioristic sense,
an argument will be advanced to show that moral behavior is
an appropriate goal of medical ethics teaching in the sense
of an ultimate or ideal goal (a goal$_u$) but not appropriate
in the sense of a proximate educational goal (a goal$_e$).
Goals$_e$ are typically identified with evaluative criteria,
and they are presumed to be constitutive or predictive of
goals$_u$. The argument aims to show that success in terms of

the Wilson+1 goals (or some other appropriate set of proximate educational goals) is all that can reasonably be demanded, but that this is consistent with adopting moral behavior as the goal$_u$ of medical ethics teaching. The argument should remove philosophers' misgivings about endorsing moral behavior as a goal and, in the process, show that the evaluation of medical ethics teaching can compare favorably with the evaluation of teaching generally.

Eschewing a Behavioristic Criterion. Behaviorists define (i.e., purport to define) behavior as "overt" and intersubjectively observable; references to agents' reasons, intentions, or mental processes are excluded. The theoretical inadequacies of behaviorism will be discussed in detail in Chapter III. It is sufficient for present purposes to observe that such a definition of 'moral behavior' effectively precludes reasoned moral evaluation. The behavioristic sense of "overt" behavior requires, in Callahan's words, "a preestablished blueprint of what will count as acceptable moral behavior" (1980, p. 70). Such a "blueprint" entails a list of the "overt" behaviors which teaching should produce, requiring a kind of unanimity on ethical issues that is difficult to imagine and inappropriate to demand. Closely related to this, the behavioristic blueprint ignores the importance of the intellectual dispositions of being critical, flexible, and sensitive to alternatives (i.e., the dispositions included in Reasoning). Such mental dispositions are central to moral

behavior and moral evaluation. Unless some obvious and serious breach of morality is involved, the appropriateness of alternative courses of action may not be decided until the particulars are known. Even then there may be room for reasonable disagreement--consider abortion, for instance-- in which case no clearly correct "overt" behavior exists.

Medical ethics teaching and its evaluation should reflect the complex and provisional nature of what counts as appropriate moral behavior, and leave room for differences among individuals on hard cases. Because the restrictive, behavioristic sense of 'moral behavior' (with its implicit requirement for a "blueprint") is inadequate in these respects, "overt" behavior in the behavioristic sense is an inappropriate goal of medical ethics teaching.

Moral Behavior is an Ultimate Goal but not a Proximate Goal. When one moves beyond a behavioristic criterion, the issue is whether moral behavior (in its ordinary workaday sense) establishes an unreasonably demanding criterion of success for medical ethics teaching. This demand (implicit, for instance, in Sider and Clements, 1984) emerges as the conclusion of the following argument:

(1) The behavior of students in educational contexts (e.g., what they say they would do on a test, questionnaire, or in an interview) can be distinguished from what they do when faced with real situations.

(2) Behavior in the educational context (behavior$_e$) is contingently related to behavior in real life situations (behavior$_r$).

(3) Evidence for behavior$_e$ is not necessarily evidence for behavior$_r$.

(4) Moral behavior$_r$ has not been directly related to medical ethics teaching, nor has it been correlated with behavior$_e$.

(5) The test of an educational endeavor is effectiveness in producing the desired behavior$_r$

(6) Therefore, medical ethics teaching (a) has not demonstrated its effectiveness, and (b) can do so only by empirically demonstrating its positive effect on moral behavior$_r$.

Making explicit the reasoning underlying the demand to adopt moral behavior (moral behavior$_r$) as a proximate educational goal (a goal$_e$) shows the demand is unreasonable. Premise (5), required to justify the demand contained in (6b), imposes a standard that few, if any, educational activities could meet. Precious little evidence exists to establish the relationship between criteria typically used to evaluate students' behavior$_e$ and their professional performance, behavior$_r$, in any area. Empirically establishing a positive relationship between behavior$_e$ and behavior$_r$ is difficult, costly, and rarely done. In the overwhelming majority of cases, a positive relationship between behaviors$_e$ and behavior$_r$ is presumed on the grounds that (1) the content of education is necessary for the behavior ultimately desired (e.g., knowing anatomy and being a good physician) or (2) the content of education is analogous to the behavior ultimately desired (e.g., taking histories from simulated patients and taking histories from real ones). Given the presumption of one of these two kinds of relationships, desired performance in terms of behaviors$_e$

is accepted as adequate to establish the effectiveness of an educational endeavor.

No arguments have been produced to show why ethics teaching should be required to meet stiffer standards. Moral behavior$_r$ may be associated with the goal$_u$ of medical teaching, and moral behaviors$_e$ may be identified with goals$_e$, the Wilson+1 goals. Because Wilson develops his "components of the morally educated" by analyzing moral behavior$_r$ into its constituent elements, the evidence provided by investigating the behaviors$_e$ associated with the Wilson+1 goals sanction inferences to moral behavior$_r$. (For example, findings about Knowledge fit presumption (1), and findings about Reasoning fit presumption (2).) Such inferences, though not unproblematic, are no more problematic than the analogous inferences made in educational research, for example, inferences about students' ability to interview real patients based on their interviews of simulated patients, or inferences about students' ability to solve real clinical problems based on their performance using paper cases.

This discussion of moral behavior as a goal ignores certain complications--most notably, the distinction between direct and indirect goals and the assumption of a relatively clean distinction between "educational" and "real" contexts unlikely to apply to much teaching in the clinical context. Nonetheless, it does establish the important conclusion that medical ethics teaching is not unique regarding behavior as

a goal of teaching. Unmasking the two senses of 'goal'—
ultimate and proximate—leads to the conclusion that the
demand to establish positive relationships between moral
behavior (moral behavior$_r$) and the effects of teaching
(moral behavior$_e$) should (1) take into account the nature
and limitations of educational evaluation, and (2) be no
more stringent than is customary for other educational
pursuits (Archambault, 1975). Given the nature of the
Wilson+1 goals—constituents of "morally educated"
behavior—medical ethics teaching that achieves them
compares well to many other educational endeavors on the
question of the relationship between performance in
educational and performance in real contexts.

In conclusion, moral behavior is an appropriate ultimate
goal of medical ethics teaching but is not an appropriate
proximate goal or evaluative criterion. Instead, proximate
educational goals, namely, the Wilson+1 goals, function as
evaluative criteria and sanction inferences to the ultimate
goal of moral behavior.

Appreciation as a Direct Goal

A controversy related somewhat to moral behavior as a
goal concerns Appreciation as a direct goal. A likely
objection is that Appreciation should be at most an indirect
goal, contrary to the way it was classified earlier. Three
reasons might be advanced. First, like the indirect goals
(e.g., Courage), Appreciation is a function of attitudes and
other causal factors largely beyond the control of

instructors. Second, the real aim of medical ethics teaching should be cognitive learning, not pleasing students. Third, students are not competent judges of what is valuable (i.e., worthy of appreciation) in their educations.

Appreciation can be defended against the first argument by invoking the <u>all things equal</u> clause presupposed in making the distinction between direct and indirect goals. <u>All things equal</u>, medical ethics instructors are in a better position to instill Appreciation than an indirect goal such as Courage. Furthermore, Appreciation is a more customary and natural outcome of teaching than progress in terms of the indirect goals.

Two answers can be given to the claim that direct aims of teaching should be confined to cognitive goals. On the one hand, empirical evidence shows that Appreciation (as measured by student course evaluations) correlates with cognitive learning (Scriven, 1981). Thus, Appreciation provides one indicator of success at achieving cognitive aims. On the other hand, confining medical ethics teaching solely to cognitive aims is too restrictive. It would render ethics teaching sterile and of questionable relevance to health care professionals. In this vein, understanding and Appreciation are not as easily separated in ethics as they might be in certain other subjects. For example, an engineering student might fail to Appreciate (i.e., fail to see the point of) algebraic derivations, even detest doing

them, and still understand them in the sense of being able to perform them when the need arises. Goods "external" to the practice of algebraic derivation (say, the satisfaction of designing a bridge and getting a fat paycheck) may motivate the engineering student. A parallel is lacking when it comes to the "practice" of ethical inquiry. Attempting to avoid malpractice suits is a reason (i.e., an "external" one) for telling patients the truth, but it is surely not a moral one. Students who do not see the point of ethical inquiry cannot be said to understand ethical inquiry, and will have no "external" motivation for engaging in it once out of the classroom. Thus, instilling Appreciation (getting students to see the point) is required in medical ethics teaching.

The claim that students are incompetent judges of what is valuable in their education is based on the paternalistic stance that students are not in a position to judge what their profession will demand and thus not in a position to judge what the content of their education should be. According to this argument, whether students Appreciate medical ethics when it is taught is therefore beside the point.

The response to this third criticism of Appreciation as a direct goal also has two aspects. On the one hand, research suggests that Appreciation (again, as measured by student course evaluations) is stable over time. That is, there is a high relationship between evaluations of courses

at the time they are taught and later retrospective evaluations (Aleamoni, 1981). More significantly, if Appreciation is not cultivated at the time ethics teaching is done, when medical and nursing students are developing their professional identities, it would seem unlikely that it will be developed later on.

In summary, although Appreciation resembles the indirect goals more than the direct goals in some respects (e.g., it is not taught directly and is not a criterion of students' performance), all things equal, Appreciation is under reasonable control of ethics instructors, is a customary educational aim, and is associated with the value of the cognitive goals of ethics teaching. Thus, Appreciation may be legitimately construed as a direct goal.

Medical Ethics Teaching and "Formalism"

A third controversy involves a recent and relatively fundamental criticism of mainstream medical ethics. Nobel (1982) has criticized medical ethics for being "acontextual"; Caplan (1983) has criticized it for employing the "engineering model"; and Clements and Sider (1983) have criticized it for being "formal". The general complaint expressed by each is that medical ethics, as presently practiced, is far removed from real ethical concerns and amounts to little more than a pointless philosophical exercise. Although the primary target is the literature of medical ethics, the criticism applies to teaching as well since discussions from the literature of medical ethics

often comprise the methods and content of teaching.

If there are such "formalists", who attempt to apply the tools of philosophical ethics with no regard for psychological, social, and historical contingencies, then they deserve to be criticized. But such individuals are rare, or at least the tide has turned against them. Influential philosophers such as John Rawls (1971) Alasdair MacIntyre (1981), Richard Rorty (1979 and 1982b), Hilary Putnam (1983), and Sissela Bok (1978) explicitly reject the "formal" approach. In one way or another, they all agree with Putnam's claim that, when applied to actual moral problems, "traditional" ethical theories (he excludes Rawls) "prove too much" and therefore prove nothing, or with Bok's claim that

> A system of moral philosophy put to such uses is like a magician's hat--almost any thing can be pulled out of it, wafted about, let fly. No one can be quite sure it was not in the hat all along. And the philosopher is often in the end his own most amazed spectator. He may not know how he did it--but the doves are aloft, the silk scarves in his hands! (p. 57)

Singer (1982), Wikler (1982),[6] and Beauchamp (1982) speak directly to the critics on the issue of how, if not by the application of ethical theories, philosophy and philosophers have any constructive contribution to make. Their answer is that philosophers contribute, not by the application of some specialized philosophical knowledge, but by consistently demanding things which characterize the discipline of philosophy, such as high standards of argument and conceptual clarity. They do not contend that

philosophers have a corner on even these, but that "applied ethicists" (most of whom are philosophers) simply spend a good deal of their time working through the problems of interest. A procedure of the kind described by Nobel--in which "moral problems must be abstracted from their social settings to appear purely moral" in order that philosophers may apply their "methods or moral reasoning" (1982, p.8)--is clearly not endorsed, at least not by the present spokespersons in the field. More to the point, present medical ethics teaching--embodied in the Wilson+1, DeCamp, and Callahan goals--is not merely formal.

Medical Ethics Teaching and Ethical Theory

    The final controversy involves the content and teaching methods naturally associated with the Wilson+1 goals. Two criticisms have been advanced by Troyer (1982). The first (something like the flip-side of the charge of "formalism") is that medical ethics teaching pays inadequate attention to ethical theory. In particular, he criticizes Benjamin and Curtis (1981) for relying too heavily on discussions of cases at the expense of ethical theory. Benjamin and Curtis do consider ethical theory and adopt "wide reflective equilibrium" (see note 6) as a general theoretical stance. But they by no means stress ethical theory. Indeed, the chapter of their text Ethics in Nursing devoted to the issue is entitled "Unavoidable Topics in Ethical Theory", much to Troyer's dismay.

Troyer's concern seems misplaced. Although some exposure to issues in ethical theory is appropriate--naive utilitarian reasoning and the problem of justice, for instance[7]--ethical theory should take a backseat to working through specific moral problems endemic to the health professions and which impinge on the concerns and interests of students in professional training. Ethical theories, after all, have historically been concerned with investigating the epistemological foundations of moral knowledge, not with application. Indeed, the test of such theories is often how well they square with moral judgments that are clear on independent, intuitive grounds. When this observation is combined with the fact that any reasonable candidate for acceptance is going to have to square with almost all intuitive judgments, it is evident why Bok would wonder whether the solution to a moral problem wasn't there in the "hat all along" and independent of the moral theory being employed. Moreover, ethical theories break down at just those places where intuitive moral judgments do. The point of medical ethics teaching ought to be to get students to make good moral judgments in the first place. For those with an interest, ethical theory can come later.

Perhaps Troyer's demand is related to his second criticism of Benjamin and Curtis: they fail to press students to come up with the "correct" position. Now, one doesn't have to view ethics as all just a matter of subjective opinion or anything of that sort to agree with

Benjamin and Curtis. In the first place, there simply may
be no good solution at hand. Perhaps a problem is truly
insoluble, or perhaps more time for thought and information
gathering is needed. In the second place, our pluralistic
society allows persons a range of freedom in their moral
beliefs. Would Troyer fail a student for insisting that
abortion is permissible? Would he fail one for insisting
that it is not? (How students are evaluated carries a
powerful message.) More to the point of ethical theory, it
is by no means clear how studying it could help with the
problem of "correct" answers. On the contrary,
disagreements between competing ethical theories have proven
intractable, and the practice of emphasizing these
disagreements in ethics teaching has led MacIntyre to
remark, "It is no wonder that the teaching of ethics is
often destructive and skeptical in its effects upon the
minds of those who are taught" (1981, p. 112). It is not
uncommon for students to declare their allegiance to an
ethical theory and then proceed to turn the crank (or play
the instructor's game), as if the declaration insulated them
from glaring difficulties in the position they wish to
defend. When the details of ethical theory are made
secondary to working through real-life issues of concern,
the likelihood is greater that "correct" positions will
emerge because it is less likely that ethical inquiry will
be viewed as an arcane intellectual exercise.

Misconceptions

The discussion now turns from "controversies" to
"misconceptions". As stated previously, the label
'misconception' applies to criticisms of a kind likely to be
voiced by students, medical school faculty, and others
unfamiliar with the nature and goals of ethics teaching.
While the controversies raised doubts about the Wilson+1
goals in particular, the misconceptions question medical
ethics teaching generally.

Ethics is Subsumed by the Social Sciences

One misconception about medical ethics teaching is
exemplified by the tendency of medical schools to lump
ethics with other disciplines, primarily the behavioral
sciences, under the portion of the curriculum designated as
"psychosocial." The mistake involved in identifying ethics
with the social sciences is related to the charge of
"formalism" in that both muddy (or deny) a distinction
between factual knowledge on the one hand and the use of
such knowledge in making moral evaluations on the other.

For the Wilson+1 goals, social scientific knowledge is
important for only one of the six characteristics (i.e.,
Knowledge) required of moral problem-solving. Although the
findings of the social sciences are relevant to
intelligently working through moral problems, the role of
such findings is limited. The social sciences, after all,
only tell what people do do, not what they ought to do. For
example, it does not follow that because physicians <u>tend</u> not

to communicate the costs of alternative treatment plans to patients in obtaining informed consent that they ought not to.[6] In short, teaching the findings of social science is not a substitute for teaching ethics. Nor does teaching the findings of social science permit ethics to take care of itself.

Ethics is Just Interpersonal Skills

A second misconception fairly prominent in medical education is that ethics can be identified with interpersonal skills. Training in interpersonal skills has become commonplace in medical education in recent years (Kahn et al., 1979), and this development is desirable except to the extent that it is believed to constitute all of medical ethics teaching. Having good Interpersonal Skills (a good "bedside manner") is important in its own right, but, like knowledge in the social sciences, Interpersonal Skills are only one element of moral problem-solving. As Margaret's dilemma illustrates, Interpersonal Skills leave a lot of territory uncovered.

Ethical Codes are Sufficient

A third misconception is that professional ethical codes provide a sufficient means for resolving ethical problems. Ethical codes are keyed to the specific problems which are endemic to various professions; they highlight these and establish ideals to which professionals should aspire. But they are inadequate to the task of giving answers to many actual moral problems. For example, the medical profession

has a long history, beginning with the Hippocratic Oath, of

protecting patient confidentially. Section 9 of the AMA

code reads as follows:

> A physician may not reveal the confidences entrusted to him in
> the course of medical attendance, or the deficiencies he may
> observe in the character of patients, unless he is required to
> do so by law or unless it becomes necessary in order to protect
> the welfare of the individual or of the community. (Veatch,
> 1977, p. 355)

The problem with this section of the code as a means by

which to resolve problems involving patient confidentially

is that it is silent on just what such problems normally

turn on: balancing the welfare of the individual against the

welfare of the community. What should a physician do when

faced with an alcoholic bus driver or a psychiatric patient

threatening to kill his girlfriend?[7] The code is of little

help.

This limitation of ethical codes is not confined to the

example in question. In general, ethical codes are at once

too broad and too narrow. They are too broad because, if

they are to carry any weight at all, they have to be

sufficiently open-ended to be acceptable to a wide variety

of viewpoints (Benjamin and Curtis, 1981). They are too

narrow because they cannot be formulated to anticipate all

of the relevant contingencies and all of the different moral

problems that might arise.

Legal Considerations Preclude Ethical Ones

Another misconception about medical ethics teaching, not uncommon among physicians, is that legal considerations preclude ethical considerations. Such a view is sometimes used to urge the futility of teaching medical ethics, and it involves several confusions. First, although individuals have a prima facie obligation to obey the law, this clearly can be overridden, and when this might be appropriate is itself a moral issue.[8] Second, there is significant overlap between law and morality. The slogan, "You can't legislate morality" is clearly false if it is taken to mean that the law does not embody moral principles. It is thus quite misleading to claim that legal considerations preclude ethical ones. Finally, like ethical codes, the law remains both too broad and too narrow. The law requires interpretation, and moral considerations frequently figure in.[9]

Ethics Teaching Conflicts with Religious Belief

Religious training invariably includes moral training. Thus there is an understandable tendency to identify religion and morality and to view secular moral education as an illicit encroachment on religious belief. Much of the force of this view is based on the fact of moral disagreement. For example, the religiously-based view that a fetus is a person, fully endowed with the right to life, cannot be reconciled with secular views which deny this. Accordingly, so the argument goes, religious and secular

moral judgment are clearly at odds: one depends on the dictates of wholly human judgment and the other on the dictates of God.[10]

Two observations count against this interpretation of the above disagreement about abortion. First, the belief that fetuses have a right to life is not necessarily religiously based. John Noonan (1977), for instance, assumes that newborns have a right to life and, without invoking any religious premises, argues that newborns cannot be distinguished from fetuses in any morally relevant ways. Second, and more generally, different moral beliefs, or "rules" do not entail different methods of moral problem solving or different basic moral "principles" (Singer, 1970). For instance, in the abortion example the moral principle 'Preserve innocent human life' is not at issue; the disagreement is rooted instead in competing views over the status of fetuses. A position on the status of fetuses in turn leads to a position on whether they should be granted the right to life. This disagreement between a secular and religious moral position, an apparently intractable one, entails neither differences about the nature of moral judgment nor that the goals for medical ethics teaching outlined are biased against religious ethics. The disagreement instead turns on opposing metaphysical premises that may or may not be religiously based.

The appeal to religion does not provide a means to put

ethical inquiry on "automatic pilot" (Benjamin and Curtis, 1981). Religious moral codes, like legal and professional codes, cannot be devised so as to anticipate all or even most eventualities. This is especially true with medical-ethical issues engendered by the advance of technology. For example, religious authorities currently grapple with concepts such as "death with dignity", "quality of life", "ordinary" versus "extraordinary" care, "proportionate" versus "disproportionate" care, and so on, in order to come to grips with current problems surrounding withdrawing treatment from the hopelessly ill. There seems no escape from giving careful attention to individual moral problems as they arise. After all, substantial disagreement over moral issues occurs within religions as well as between them; there is nothing even approaching a united religious front.

Social cooperation in general would be impossible without considerable agreement on most issues. General agreement does exist, even among cultures, and this fact is difficult to account for if there are not shared moral standards that cut across religious boundaries (Nowell-Smith, 1967). Where specific disagreement does exist in our own culture, it is in the interests of persons in general to use secular moral argument to win assent to their views on controversial issues. Anyone who endorses freedom of religion would be hard pressed for a viable alternative.

Ethics Teaching is Indoctrinating

Another general misconception about ethics teaching is that training in ethics entails indoctrination. The conditions under which an educational endeavor amounts to indoctrination raises some rather thorny philosophical issues that cannot be considered here. (See Macklin, 1980, for a recent review.) In its simplest terms, the issue boils down to whether any reasonably "neutral" approach is possible. Provided that teaching is balanced in the views that it presents, the fear of indoctrination is unjustified.

The issue of indoctrination can be turned on its head. That is, the consequences of _not_ engaging students in ethical inquiry should be considered. Such a "hands off" approach only reinforces the all too prominent belief that ethics is simply a matter of personal preference. One must also wonder whether this approach does not actually increase the likelihood of indoctrination by failing to provide students with the skills necessary for fending off would-be indoctrinators.

Formal Education has no Proper Role to Play in Moral Development

The final misconception is that moral development is not a proper function of formal education. The misconception finds its roots in subjectivism--a view of ethics that conflates rational justifications for moral beliefs and the causes of such beliefs, and which maintains a strict dividing line between facts and values. Subjectivists see no roles for rationality and formal education in connection

with ethics. Rationality and formal education apply only to knowledge and science, and exclude the preferences and feelings which subjectivists believe exhaust ethics. Given this view, the most that can be done by way of medical ethics teaching is the inculcation of the conventional standards of the profession.

If MacIntyre (1981) is correct, subjectivism ("emotivism") characterizes our culture in general and pervades social research in particular. If unchecked, this poses a fundamental obstacle to the evaluation of medical ethics teaching as well as teaching ethics per se. Removing this obstacle will be one of the central aims of the next chapter.

Conclusion

The following seven goals have been described, defended, and proposed as a framework for the evaluation of medical ethics teaching:

Direct Goals:	1. imparting Knowledge
	2. improving Reasoning
	3. instilling Appreciation
Indirect Goals:	4. stimulating Moral Regard
	5. eliciting Empathy
	6. reinforcing Interpersonal Skills
	7. promoting Courage

Key issues in their development were representativeness, defensibility against criticisms, and the distinction between direct and indirect goals.

The central aim of this chapter was to lay a foundation for the credibility of the evaluation of medical ethics teaching by establishing the credibility of a framework of goals. This in turn required demonstrating the representativeness of the goals and their defensibility against criticisms. Representativeness of the goals, required for them to be received as relevant by medical ethics teachers, is demonstrated by their consistency with those advanced by experts in medical ethics, namely, the Hastings Center and DeCamp groups. Defensibility of the goals, required for them to be received as other than self-serving by interested groups in general, is demonstrated by their ability to withstand criticism from both inside and outside the field of medical ethics.

The distinction between direct and indirect goals is based on the moral and practical differences between the sets when viewed in the context of professional education. Moral behavior is the ultimate aim of medical ethics teaching. Its complex nature, reasonable expectations regarding the impact of education, and general educational evaluation practices, however, preclude adopting it as a proximate goal. Instead, surrogate constituents of moral behavior, i.e., the Wilson+1 goals, comprise appropriate proximate goals and hence appropriate evaluative criteria.

Given this brief summary, three further clarifications about goals are in order. First, although the Wilson+1 goals are representative, they are nevertheless

prescriptive. That is, they are goals to be urged and used in addition to or in the place of goals which may be avowed in particular contexts (though justifiable adaptations and improvements are always appropriate).

Second, the seven goals comprise a minimal set for medical ethics programs. That is, an adequate program ideally would have to give some attention to each. Individual activities (e.g., courses or lectures), however, might legitimately focus on some goals to the exclusion of others. In this connection, relatively small programs or new and inchoate ones might legitimately give little attention to indirect goals in response to the need to order priorities. Additionally, the general milieu of professional education creates an obstacle for any educational goals that may be viewed as promoting moral virtues (like Courage). Although there are demands to demonstrate effects in terms of moral behavior, there are also marked countervailing demands to avoid advocating or judging the shape such behavior should take.

Finally, the Wilson+1 goals are relatively silent on the issue of curricular "content" (i.e., the issues and concepts which should be taught). Although this is an important issue for evaluation, (Callahan and the DeCamp group include it in their statements) it is beyond the scope of this study.

NOTES

1. The terms Moral Regard, Empathy, Interpersonal Skills, Knowledge, Reasoning, and COURAGE are mine, not Wilson's. He uses PHIL, EMP, GIG_1, GIG_2, DIK, and KRAT respectively. He also includes a component omitted by me, namely, PHRON, which corresponds roughly to personal prudence. In later writings (1969 and 1973), Wilson refines and further explicates his components. These later versions are unwieldy, at least for my purposes.

2. From Benjamin and Curtis (1981, p. 162).

3. The conferees were Charles Culver, Dan Clouser, Bernard Gert, Howard Brody, John Fletcher, Al Jonsen, Loretta Kopelman, Joanne Lynne, Mark Siegler, and Dan Wikler. The results of this conference are now available in published form (see Culver et al., 1985).

4. Wikler challenges Nobel for a positive account of moral epistemology, and makes reference to Daniels' (1979) discussion of the Rawlsian notion of "wide reflective equilibrium" as one alternative to her position (or (non-position) on this matter. Given Daniels' account--one which squares with present day anti-foundationalist (or anti-formalist) views--Nobel's charges of "acontextualness" and "abstracting the purely moral" do not apply. The account makes explicit reference to the import of empirical knowledge as well as common moral intuitions.

5. Naive utilitarianism is the unqualified view that an action is right if and only if it promotes the greatest good for the greatest number (the principle employed by Roskolnikov to justify robbing and murdering the pawn broker). The conflict between the utilitarian principle and justice is glaring. For example, it sanctions using the indigent (the most vulnerable group) in medical research with no regard for informed consent provided only that the good which will be derived outweighs whatever harm is done to the experimental subjects. Naive utilitarianism is insidious because the more vulnerable the group being unjustly treated the less likely there is to be an outcry and the _easier_ it is to achieve a positive balance of good over harm.

6. Some 38% of physicians do not believe it is necessary to inform patients of the costs of therapies and procedures; whereas, 70% of patients believe such information is relevant to deliberations (The President's Commission, 1982). This is evidence of a "performance deficit" (a failure to perform appropriately) on the part of physicians, not a guide to moral behavior.

7. _Tarasoff v. Regents of the University of California_ (Beauchamp and Walters, 1978, pp. 176-185) is a landmark case which established that the confidentiality of the doctor-patient relationship is not protected absolutely. The court ruled that physicians, particularly psychiatrists, have a "duty to warn" individuals whose lives may be in danger, even if the information is obtained in contexts believed to be confidential by patients who pose the threat.

8. Acting contrary to the law for moral reasons is the basis of civil disobedience, which is generally viewed as laudable. In the context of the practice of medicine, one can imagine a physician who periodically failed to comply with reporting requirements for child abuse and neglect laws on the grounds that reports result in greater harm than benefits to children in certain instances.

9. See Hart (1961) for his discussions of the "open texture" of the law (the feature which makes a continuing need for interpretation unavoidable) and of the relationship between law and morality.

10. The notion that morality is and must be wholly grounded in commands issued from God (or the gods) was first criticized by Plato in the _Euthyphro_. Wilson exhibits a view similar to Plato's in the following:

> Do they [certain religious persons] really want to maintain that any religious or non religious authority can _in itself_ act as an acceptable basis for morality: that commandments must be obeyed _just_ because they are commandments? Do they rather not believe that God (or the Church, or the State, or whatever authority it may be) ought to be obeyed because he says what is right or good, rather than _just_ because he is God? And, if so, will not what the authority says about morality turn out to be justifiable in terms of our own criteria [i.e., the components], so that they need have nothing to fear from their application? Do they not think they are _justified_ in accepting the authority--that they accept it for some _good reasons_, rather than as a completely wild leap in the dark, a leap which (unless backed by some kind of good reason), would be logically indistinguishable from leaps made by Fascists, Baal-worshippers, or lunatics? And if all this is so, can we not find common ground, at least so far as morality is concerned, on the basis of those reasons for morality which we all accept as primary? (1969, p. 180)

CHAPTER III

BEYOND SCIENTISTIC EVALUATION:

DEFENDING A FUNCTIONAL APPROACH

> Education...is not an enterprise which we know just how to
> handle and can reduce to a series of techniques, like
> vine-pruning or cutting corn or building bridges. It is a much
> more general enterprise--like marriage, religious counseling,
> childrearing, and psychotherapy. Such enterprises are shot
> through and through with conceptual confusion and uncertainty,
> and the necessity of making value judgments (Wilson, 1983,
> p.192).

Evaluation is not "an enterprise which we know just how

to handle and can reduce to a series of techniques"

either--it, too, is "shot through and through with

conceptual confusion and uncertainty and the necessity of

making value judgments."

The preceding chapter clarified _what_ to look at; this

chapter will clarify _how_ to look at it. This will be a

complex task because many evaluators in fact endorse a

"vine-pruning" approach: "Designs too often reflect a narrow

operationalism based on obeisance to the controlled

experiment as the design ideal" (Patton, 1983, p. 29). An

attenuated form of positivism predominates in educational

evaluation, inherited from behavioristic educational and

social science research more generally. A rigid fact-value

distinction, a related distinction between quantitative and

qualitative methods, and a presumption that measurement must

be grounded in psychological theory are central features of what shall henceforth be referred to as "scientistic"[1] evaluation.

Scientistic evaluation is flawed in general but is especially obstructive to the evaluation of medical ethics teaching. Given a fact-value distinction that places values beyond rational scrutiny, evaluating teaching in the realm of moral values becomes an incoherent activity. If morality is simply a matter of tastes, feelings, and upbringing, and all positions are equally rational, then there are no criteria to support evaluative judgments about the performance of students.[2] Given a quantitative-qualitative distinction, which labels qualitative methods unscientific and impressionistic, the evaluation of ethics teaching is bound to be labeled "soft". All that is relevant to the evaluation of ethics teaching cannot be reduced to a collection of operationalized characteristics, measured with paper-pencil tests and checklists of <u>overt behavior</u>, and analyzed statistically. Qualitative methods, such as interviews and direct observation, are indispensable. Finally, the presumed superiority of criteria and measures grounded in psychology leads to an emphasis on reliable measures at the expense of valid ones[3] that Messick (1981) has dubbed the "tyranny of reliability". In connection with the evaluation of ethics teaching, evaluators search the terrain for some proven instrument--proven in the sense that it measures reliably--with little or no attention paid to

<u>what</u> is being measured. As a result, measures of attitudes, overt behavior, and Kohlberg's stages of moral development have been frequently advocated or used in the evaluation of medical ethics teaching with far too little attention paid to the validity of such measures for evaluating teaching.

This chapter criticizes each of the three features of scientistic evaluation—its construal of the fact-value and qualitative-qualitative distinctions and its presumption in favor of criteria and measures grounded in psychological theory—with an eye toward vindicating an alternative "functional" evaluation approach. The functional alternative holds that evaluation should emphasize adjusting methods and measures to fit different objects and circumstances of evaluation (Cronbach, 1982; Cronbach and Associates, 1980; and Patton 1980); it holds that the relevance and usefulness of information, not a priori judgments based on abstracted criteria of methodological rigor, determine the quality of evaluation.

The Fact-Value Distinction

Persons who take the evaluation of ethics teaching seriously too frequently encounter the following general question: How can you <u>evaluate</u> ethics teaching? Seemingly innocent, this question poses a fundamental challenge to the evaluation of ethics teaching because it is so often motivated by ethical subjectivism, i.e., it is so often motivated by the belief that there is nothing to be "right"

about in ethics and thus nothing to evaluate.

Ethical subjectivism is an especially serious obstacle when it is exhibited in groups such as teachers and those responsible for evaluation, and subjectivism is indeed prevalent among these groups. In addition to MacIntyre's observation about the pervasiveness of ethical subjectivism in social research mentioned earlier, Scriven claims that to deny that value judgments are essentially subjective and undecidable "even in these days of the decline of positivism requires considerable intestinal fortitude and intellectual competence" (1979, p.13). In response to current thinking among evaluators and the "value-free doctrine", he remarks,

> ...attacking the value-free doctrine accurately and effectively is important because there are so many spurious reasons for adopting it and so much value-phobic pressure to accept it...if the doctrine is not rendered completely absurd by complete exposure, it will simply continue to rise from the ashes (1983, p.81).

One doesn't have to look too far to find instances of Scriven's "value-phobia". Anderson and Ball (1978) devote three chapters of The Profession and Practice of Program Evaluation, a standard text, to ethics and values in evaluation. The following sample of passages is eye opening:

> [1] Given the effect of ideology on evaluation, what is the evaluator's professional obligation? What should the evaluator do to inform or, perhaps, protect the program director and the other audiences for the evaluation? ...We should emphasize two principles--the evaluator cannot become a Spock-like emotionless Vulcan, and it is worthwhile for the evaluator to make explicit, in as honest and open a way as possible, the values he holds (p. 115).

[2] The evaluator had seen what the evaluator had wanted to see. (Or maybe the evaluator was correct in his perceptions, and our perceptions were biased by our distaste for such a subjective procedure) (p. 118).

[3] Messick...discusses the impact of these professional values in an excellent paper. He points out that values pervade not only our decisions on where to look but also our conclusions about what we have seen. And he presents an illustration that is better left in his words than paraphrased in ours (lest our values distort his meaning!) (p. 118).

[4] Evaluators with different basic social, moral, and economic values, different predispositions, and different preferences will perform different evaluations. Unless the audiences for the evaluations are informed about these underlying influences misconceptions are likely to arise (p. 124).

[5] To the extent possible and without creating intolerable confusion, the evaluator should inform audiences for the evaluation results how...results are based upon a particular evaluation approach...The evaluator will generally have to depend on a simple accounting of what decisions were made during the evaluation, why those decisions were made, and what the major alternatives were. This is the honest and open approach. Given our values, we recommend it! (p. 125)

Viewed in isolation from one another, the first four of these passages may appear reasonable and, indeed, true. But together they paint a rather clear picture which the last passage (the closing one of the chapter) brings into sharp focus. The message is that there is an obligation to get values out on the table, but no suggestion is made about what ought to be done with these values once they are made explicit.[4] What is worse, it looks as if value claims are always up for grabs and are to be made only with great trepidation. The qualifier in the last passage, "given our values", brings home the value phobia--as if one had to hedge on the demand for an honest disclosure of relevant information and had no justification for writing-off autocrats.

The value phobic fact-value distinction has at least two sources: positivistic epistemology and the attempt to avoid bias. Donald Campbell appeals to both reasons, and at a level of philosophical sophistication unusually high among educational researchers. His views are well-suited to frame a general examination of the issues. The conclusions of this examination of Campbell's views may then be related to the evaluation of medical ethics teaching in particular.

The Positivistic Justification

Campbell follows the logical positivists on the fact-value distinction by his own admission; his basic epistemological stance is contained in the following:

> The tools of descriptive science and formal logic can help us implement values which we already accept or have chosen, but they are not constitutive of those values. Ultimate values are accepted but not justified (1982, p. 123).

The claim that "ultimate values are accepted but not justified" is true in the sense that not all values can be up for grabs at once, but the contrast Campbell presumes between values and the "tools of descriptive science and formal logic" does not exist. Campbell himself elsewhere (1974) argues for the "presumptive" nature of scientific knowledge. That is, he shares the general Kuhnian view that knowledge is based on theory-laden observations and beliefs. The consequence of construing knowledge as presumptive, which Campbell fails to appreciate, is that there can be no "ultimate" justification for beliefs of any kind, including basic scientific ones. As Wittgenstein observes, "Justification...comes to an end. If it did not it would

not be justification" (1958, p. 136e). Wittgenstein's observation applies as well to scientific claims as to value claims. Thus, Campbell's way of distinguishing factual and value claims--on the basis of whether they must be "accepted but not justified"--fails.

The positivistic fact-value distinction is, after all, based on positivism's central (and contra-Kuhnian) notion that purely observational, atheoretical knowledge can be isolated and used as the basis for theory. To qualify as a legitimate knowledge claim, to be "cognitively significant", a sentence had to be testable in one of two ways: either in terms of direct observation or in terms of formal logic. (Notice the similarity to Campbell's "tools of descriptive science and formal logic".) Since value claims could not be verified (or falsified) in either of these ways, they were judged to be devoid of cognitive content, to be mere expressions of emotions. The positivists more or less backed into "emotivism"--a view that equates 'Abortion is wrong' with evincing an emotion, e.g., 'Boo abortion!'--as a result of their more general epistemological position. The often overlooked but crucially important implication of the abandonment of positivism is this: If the positivistic attempt to ground all knowledge in some sort of atheoretical reality is untenable (which few, including Campbell, nowadays deny), then the justification for the rigid distinction between facts and values is equally untenable. The positivistic construal of the fact-value distinction is

merely a corollary of the more general observation-theory distinction.

Because the positivistic fact-value distinction is baseless, it is illicit to separate value judgments from the conduct of research, especially social research, on the grounds that values are merely emotive and non-cognitive. In his discussion of post-positivistic social science, Scriven puts the matter as follows:

> There is no "ultimate observation language..." Analogously, there is no ultimate factual language. And the more interesting side of this coin is that many statements which in one context clearly would be evaluational are, in another, clearly factual. Obvious examples include judgments of intelligence and of the merit of performances such as those of runners of the Olympic Games (1969, p. 199).

Scriven concludes, "...there is no possibility that the social sciences can be free either of value claims in general or of moral value claims in particular..." (p. 201).

Richard Rorty echos Scriven:

> ...there is no way to prevent anybody using any term "evaluatively." If you ask somebody whether he is using "repression" or "primitive" or "working class" normatively or descriptively, he might be able to answer in the case of a given statement, made on a given occasion. But if you ask him whether he uses the term only when he is describing, only when he is engaging in moral reflection, or both, the answer is almost always going to be "both". Further—and this is the crucial point—unless the answer is "both," it is not the sort of term that will do us much good in social science (1982, p. 195-96).

Scriven's and Rorty's conclusions follow from the refutation of the central tenets of positivism. These tenets will be explored further in the subsequent section on the quantitative-qualitative distinction. It is sufficient to observe at this point that ultimate, theory-free, factual

knowledge cannot exist, and that the corollary fact-value distinction is untenable. As a consequence, setting values to one side because they don't have to do with matters of fact is indefensible. Value judgments are subject to rational criticism, and no researcher can avoid value commitments (whether such commitments are acknowledged or not). It is absurd, for example, to suggest that the participants in the Manhattan Project were engaged in a value-free enterprise, i.e, one not subject to moral evaluation. Value commitments are even more fundamental to the conduct of social research, because, as Scriven and Rorty observe, the very concepts that social researchers employ are evaluative of human behavior.

Avoiding Bias

An epistemological justification for the fact-value distinction went the way of positivism, but the distinction persists--rises from the ashes a la Scriven--as the result of misguided efforts to safeguard pluralism and avoid bias. For example, consider the motives Campbell (1982) exhibits in the following:

> An established power structure with the ability to employ applied social scientists, the machinery of social science, and control over the means of dissemination produces an unfair status quo bias in the mass production of belief assertions from the applied social sciences...This state of affairs is one which...I deplore, but I find myself best able to express my

> disapproval through retaining the old-fashioned construct of truth, warnings against individually and clique selfish distortions, and a vigorously exhorted fact-value distinction...(p. 125)

> [The] effort to make us aware of biased-paradigm co-optation is again one best done by retaining a traditional fact-value distinction; it is a matter of becoming self-critically aware of our profoundly relativistic epistemologic predicament and using this awareness in the service of a more competent effort to achieve objectivity, rather than employing it to justify giving up the goal of truth (p. 126).

Campbell clearly opposes the manipulation of social scientific knowledge in a way that serves the interests of powerful groups--a laudable position. Rather than avoiding bias, however, by waving the banner of value-free social science, Campbell is more likely contributing to bias by obscuring tacit value commitments.

To illustrate this danger, consider the concept of intelligence and whether it is possible to strictly separate truth, facts, and values. (The possibility of such a strict separation is an implication of Campbell's view and has recently been defended explicitly by Jensen, 1984). If research on intelligence involves the "goal of truth" solely, it should be possible to divest 'intelligence' of evaluative meaning. 'Intelligence', however, is one of Scriven's "obvious examples" of a concept with both descriptive and evaluative uses. This two-edged feature of 'intelligence' is not a problem per se: social science concepts must have this feature to be useful for guiding practice. That is, unlike 'velocity' qua physical science concept, 'intelligence' qua social science concept would be worthless if it did not have an evaluative use. The quest,

then, for the unvarnished, value-free "truth" about intelligence is misguided.

This quest introduces the potential for bias, and not merely the kind of bias involved in the outrage of administering intelligence tests to those not fluent in the language of the test.[5] Intelligence tests measure characteristics that are implicitly viewed as valuable or good (which, again, is what makes 'intelligence' a useful social science concept) and therefore introduce the possibility of a more fundamental kind of bias. It is not too hard to imagine a society in which "I.Q. tests" would measure the ability to construct a bark canoe. Closer to home, just as the label "intelligent" entails roughly "having something good", the label "mentally retarded" entails roughly "lacking something good." The fate of some famous "Baby Does"--anomalous newborns allowed to die for their "own good" and the good of others--is dramatic testimony to the evaluative meaning of intelligence and its consequences.[6] Less dramatic but much more common are the well-known effects of labeling school children on the criterion of intelligence (and on numerous other criteria as well).

The problem with Campbell's "vigorously exhorted fact-value distinction" is the implication that issues of value can (and should) be bracketed and set to one side while researchers go about the task of collecting purely descriptive data. At an epistemological level, because the

positivistic fact-value distinction is untenable and social science concepts are inherently evaluative, it is impossible to sharply distinguish factual from value claims. At the level of practice, the attempt to bracket values in the name of truth and science, to protect pluralism, and to avoid bias, only results in a more insidious bias. If one insists that value judgments are irrevocably non-cognitive, or biased, or to be "accepted but not justified", one is left with the conclusion that social research is flawed in the same way. The way to avoid this conclusion is to accept that value judgments are part of the fabric of social research and to recognize that they must be defended and may criticized like any other kind of judgment.

> Value judgments, like factual judgments and theoretical analyses, are of two kinds--the well-supported and the poorly supported. No scientist can avoid making them, although it is certainly possible to avoid making good ones (Scriven, 1983, p. 81)

Conclusion

As stated earlier, these general conclusions about the need and unavoidability of using value-laden concepts and of making value judgments in social research may be made specific to the evaluation of medical ethics teaching. In the preceding chapter, seven goals were proposed as a framework for evaluation. Although these goals may be open to criticism of various kinds, a criticism which does not apply is the global one that the goals are value-laden and therefore objectionable. Moral Regard, Empathy, Interpersonal Skills, and so on, are construed as good

things, but they are also descriptions. In this, they are just like other educational goals and criteria such as, achievement, cognitive skills, positive attitudes, and so on. The fact-value distinction thus cannot render the evaluation of ethics teaching suspect or untenable unless it reduces all educational evaluation (indeed all of social science) to the same position. At the most general level, the question, "How can you _evaluate_ ethics?" may be answered, "The same way you evaluate anything else".

The Quantitative-Qualitative Distinction

The identification of quantitative methods with something epistemologically respectable and qualitative methods with something epistemologically suspect is the second feature of scientistic evaluation this chapter set out to criticize. The quantitative-qualitative distinction[7], although less directly related to the evaluation of ethics teaching than the fact-value distinction, requires careful consideration because those who espouse it as marking some deep rooted epistemological distinction between the scientific method and some other non-scientific one are likely to scoff at ethics as an unscientific, second-rate pursuit. The same positivist view of knowledge which undergirds a rigid fact-value distinction undergirds a quantitative-qualitative distinction based on the purported difference between science and non-science. Thus, the fact-value and quantitative-qualitative

distinctions go hand in hand. In so far as qualitative methods are indispensable for evaluating ethics teaching, they need to be freed from their status as the "soft" and subjective handmaiden of quantitative methods, lest the evaluation of ethics teaching suffer from guilt by association.

A combination of quantitative and qualitative methods will be advanced as the appropriate methodological stance. Although advocating such a combination is not new, the practice has been questioned on the grounds that it is an ad hoc expedient and, accordingly, is epistemologically incoherent.

The positivists can claim as much responsibility for the rigid quantitative-qualitative distinction as for the rigid fact-value distinction. The advent of positivism prompted a debate over whether social research should employ the physical science model portrayed and advocated by positivism, or should employ some alternative "interpretative" model of its own. This positivist-inspired forced choice has set the terms of the contemporary debate about quantitative versus qualitative methods and--where the positivistic physical science model is identified with quantitative methods and the interpretive model with qualitative methods--limits the positions to two: one may divorce the issue of research methods from more abstract epistemological issues and employ whatever method or combination of methods seems to make sense (e.g., Reichardt and Cook, 1979), or one

may hold that abstract epistemological issues dictate methods and seek to reconcile the competing positivistic and interpretive views (e.g., Smith 1983a and 1983b).

This creates a dilemma. Reichardt and Cook offer good arguments in support of combining quantitative and qualitative methods, but their general suggestion that the two "paradigms" of research (roughly, positivistic and interpretive) are logically independent from the means of obtaining knowledge is a heavy price to pay (and they should be suspicious of "paradigms" that are independent of methods). On the other hand, Smith is correct to require a logical connection between epistemology and research methods; but, by tracing out the implications of the positivist versus interpretive frameworks, he winds up with the unwelcome conclusion that qualitative and quantitative methods "do not seem compatible given our present state of thinking" (1983a, p. 12).

The plan of this section is to escape both horns of the dilemma by criticizing the underlying positivistic presuppositions that generate it. The result will be the the best of both worlds: a free hand to use whatever method or combination makes sense and a legitimate epistemological foundation for doing so.

Qualitative and Quantitative Data

The most frequent positivist-inspired charge against qualitative data is that it is "subjective". Scriven (1972) responds that 'subjective' is ambiguous and that trading on

this ambiguity leads to erroneous conclusions about the merit of qualitative methods. He distinguishes between "quantitative" and "qualitative" subjectivity. To say that a claim is "quantitatively" subjective roughly means that it is based on the observations and arguments of a few; to say it is "qualitatively" subjective roughly means that it is highly contestable. Scriven's crucial point is that a claim that is subjective in one of these senses is not necessarily subjective in the other sense. For example, at one time the claim "The earth is round" was quantitatively subjective, but it proved not qualitatively subjective because it was backed by convincing evidence and reasoning. On the other hand, "Chocolate ice cream is better than rocky road", is qualitatively but not quantitatively subjective. Although many would assent to this claim, it is inappropriate (and unimportant) to try to establish whether it is <u>correct</u>.

Scriven's distinction provides some help in clarifying the sense in which data ought not be subjective and the sense in which whether data is subjective is beside the point. The best course, however, is to dispense with the subjective-objective distinction in discussions of research methodology precisely because of the ambiguity Scriven identifies. The real issue for epistemology and for research (and I think this is essentially the point Scriven is making) is fallibility: to disparage qualitative data as subjective is to accuse it of having high fallibility (H-fallibility); to laud the objectivity of quantitative

data is to construe it as having low fallibility (L-fallibility).

For the positivists, the limit of L-fallibility for empirical claims (an attempt at infallibility) was atheoretical, "protocol" sentences. Concerted efforts, however, to produce a satisfactory explication of the relationship between such "purely" observational protocol sentences and scientific theories (which, if successful, would have met the positivists' goal of reducing theory to a logical concatenation of observation sentences) met with failure. The concept of "cognitive significance" (the verifiability criterion)[8] grew ever more complex and unwieldy and was eventually abandoned. As Phillips observes (1983, p. 7): "The principle of verifiability suffered the same fate as the 'Elephant Man'--it became a contorted monstrosity that choked under its own weight". (One might say, alternatively, that the poor creature was choking and Quine administered euthanasia.)

The best post-positivistic approximation to the protocol sentence (the limit of L-fallibility for empirical claims) is Quine's "observation sentence". Although it does not measure up to the demands of the positivists, it "accords with the traditional role of the observation sentence as the court of appeal of scientific theories" (1969a, p. 87). His characterization is as follows:

> An observation sentence is one which all speakers of the language give the same verdict when given the concurrent stimulation. To put the point negatively, an observation sentence is one that is not sensitive to differences in past

experiences within the speech community (1969a, pp. 86-87).

Two things to note about this definition. (1) Given the positive formulation, observations are based on the criterion of intersubjective agreement among observers and, accordingly, always retain some degree of fallibility (no matter how small) since it is possible for virtually everyone to be mistaken. (2) Given the negative formulation, observation sentences are objective (or unbiased) in that they do not depend on irrelevant idiosyncracies of observers.

As observations move away from the limit of Quine's observation sentences the issue of fallibility becomes complicated, and this is the crux of the distinction between quantitative and qualitative data. At first glance, quantitative data might appear to be uniformly superior. For example, counting the number of students in a classroom generates quantitative data. The claim 'There are x students' is an instance of Quine's limiting case of an observation sentence; the data is markedly L-fallible. By contrast, observing the workings of a classroom in terms of the group dynamics results in qualitative data far removed from the limiting case and markedly H-fallible. Given this example, quantitative data is much better than qualitative data on the criterion of fallibility, and such examples no doubt account for the tremendous faith placed in quantitative data. The essential point is that this example cannot be generalized; just the opposite ordering of

fallibility between quantitative and qualitative data is possible and indeed common.

Consider pilot testing an attitudinal instrument. The ultimate aim of developing the instrument is to gather quantitative data. Experience has taught that attitudinal measures frequently suffer from difficulties in interpretation, which may render data of questionable validity. In other words, in the absence of concerted development efforts, attitudinal measures tend to be H-fallible relative to the questions of interest. What is the solution? It is to get some subjects together, administer provisional versions of the instrument, and request their opinions about interpretation. Since their opinions constitute qualitative data, and since this data is being used to reduce the fallibility of the quantitative data to be ultimately collected, the qualitative data is presumed to be less fallible than potential quantitative data. Judgments about validity have this characteristic in general, i.e., they are undergirded by qualitative data and judgment.

To conclude the discussion of quantitative versus qualitative data, the nature of concepts used in educational research--concepts like intelligence, reasoning, achievement, and attitudes--is such that ultimate dependence on qualitative judgments and data always lies at the bottom of minimizing the fallibility of quantitative instruments. So long as educational research remains couched in terms of

such concepts (and it must to have a bearing on practice) quantified data gathering will have to remain faithful to and parasitic on qualitative judgments and terms; the latter cannot be eliminated.

This section does not aim to throw out the baby with the bathwater, i.e., the aim is not to disparage quantification. The section has been devoted to a defense of the use of qualitative data because the use of quantitative data has not been required to endure the same amount of unjustified criticism. Quantitative data is valuable because, following Campbell (1974 and 1979), quantitative data goes beyond and provides a check on qualitative data. Provided relatively L-fallible (reliable and valid) instruments are employed and (not unrelated to this) suitable research questions are addressed, quantified data has distinct advantages over qualitative data. Generally speaking, quantified instruments allow attention to be focused on variables of interest, they reduce distractions or "noise", and they permit finer discriminations. Moreover, the introduction of mathematical symbols permits an economical summary of data that in turn facilitates analysis. (Imagine trying to do arithmetic in English, with no mathematical symbols.) Comparisons on the variables of interest become manageable, and the magnitudes of differences can be investigated. Finally, once the instrument development stage is completed, quantitative data can be much more efficient. (Imagine sending out an army of observers to obtain the information

available from standardized achievement testing.)

Scientific Inference

Data, whether quantitative or qualitative, is used to support inferences, and the way the positivists construed _scientific_ inference in particular contributes markedly to engendering the forced choice between the physical science and interpretive models of social research.

The positivists' general view of scientific inference had two important features: (1) scientific inference consists of confirming (disconfirming) quantitative theories and laws by appeal to their logically inferred observational consequences, and (2) the logic of scientific inference is the same for social science as for physical science. If the positivistic construal is correct, then quantitative and qualitative methods are indeed incompatible, for the use of quantitative methods in social research would be tantamount to the attempt to build quantitative laws. Both of the positivists' claims about scientific inference were wrong. Showing how they were wrong will remove the final vestige of the rigid distinction between quantitative and qualitative methods

Inference in Physical Science. Post-positivistic

philosophy of science rejects the positivistic notion that the relationship between empirical evidence and corresponding laws and theories is a precise one, explicable in terms of formal logic. The post- positivistic (Quineian-Kuhnian) view is roughly as follows: one begins

with a hypothesis which is tested against evidence deemed
appropriate. The evidence will either provisionally confirm
the hypothesis or prove inconsistent with it. In the latter
case, the evidence may either be discounted (attributed to a
poor reading of the results, inaccurate measurement, or
simply viewed as anomalous), or it may be accepted as
falsifying. If it is accepted as falsifying, then matters
become complicated because the empirical test does not apply
to the hypothesis in isolation, but to a constellation of
beliefs (a "conceptual scheme") in which the hypothesis is
embedded. In effect, a conjunction of beliefs, including
the hypothesis of interest, is put to the test rather
than--as positivism would have it--some deductive consequent
that may be directly confirmed or falsified. When evidence
is accepted as falsifying, some further choice must be made
regarding which belief included in the conjunction is
affected, and this decision cannot be read off, as it were,
from the evidence provided by the empirical test. In
addition to the evidence from given empirical tests, other
empirical beliefs, metaphysical beliefs, and general guiding
principles--simplicity, scope, and familiarity--come into
play in deciding how the evidence should affect the shape of
the revised conceptual scheme (Quine, 1970).

In the physical sciences, quantified laws are employed
that significantly circumscribe the area of interest and
dictate what is to count as confirming and disconfirming
evidence. Nonetheless, the demise of positivism entails

that quantitative evidence, even in the physical sciences, can never be interpreted independent of extra-observational and extra-theoretical (qualitative) considerations that help define both the theory in question and the broader conceptual scheme in which the theory is embedded; quantification does not eliminate qualitative judgments and therefore is not an **alternative** to them.

Inference in Social Research. The purpose of social research is to improve human practices, and this counts against the second feature of the positivistic construal of scientific inference, namely, that the same characterization of inference applies to social research that applies to physical science. If social research is to inform constructive change, it has to employ the kind of two-edged concepts described previously. Consequently--and this is where it importantly differs from physics--the concepts used in social research must be validated in terms of human interests and practices.[9] Because quantitative data must be grounded in such two-edged concepts, only modest networks of quantitative laws are possible--concepts like 'reasoning', 'achievement' and 'attitude' do not readily lend themselves to relationships like $f = ma$.[10] Even this modest level is rare and controversial, and nothing like the physical sciences. Inferences based on quantitative social science data are therefore much more piecemeal and disjointed than inferences in physical science, which is to say they require extra-theoretical (or qualitative) judgment to a

dramatically higher degree[11].

Although the fit between hypotheses, the empirical tests associated with them, and resultant data is much looser in social research than in physics, there is no incoherence in employing quantitative methods to investigate non-nomological issues. Even in the physical sciences, where the aim _is_ quantitative law-building, quantitative findings do not dictate all of the scientific judgments that have to be made. Again, various assumptions and beliefs within both a theory itself and the broader conceptual scheme in which the theory is embedded invariably come into play.

To conclude the discussion of the quantitative-qualitative distinction, criticisms were advanced at two levels. At the level of data, qualitative data are not a priori highly fallible. Indeed, quantified measures always presuppose qualitative data and judgments. At the level of inference, the belief system in question always incorporates substantive qualitative beliefs that play an ineliminable role in drawing conclusions. The consequence is that quantitative and qualitative methods are not incompatible. On the contrary, the methods are inextricably interwoven, and all who advocate combining quantitative and qualitative methods are thus on solid epistemological ground.

Conclusion

Champions of scientistic evaluation identify quantitative methods with objective data and scientific

inference and identify qualitative methods with subjective data and non-scientific inference. They disparage functional evaluation because it permits practical research aims to determine research methods and thereby compromises methodological rigor.

The scientistic view is undermined by observing that fallibility reduction (not "objectivity" or "the scientific method" per se) is the interesting epistemological issue. Quantitative data and quantitatively-based inferences are not a priori less fallible than qualitative data and qualitatively-based inferences. The aim of evaluation is reducing fallibility about questions of interest, and the questions of interest have to do with improving human practice. Meeting this aim entails employing a two-edged (evaluative-descriptive) vocabulary that renders inference in the social sciences very unlike that in the physical sciences, straining the scientistic view to the breaking point.

The discussion of the scientistic construal of the qualitative-quantitative distinction may be concluded by expanding on an analogy attributed to Hilary Putnam: "If you want to know why a square peg does not fit into a round hole you had better not describe the peg in terms of the positions of its constituent elementary particles" (Rorty, 1982b, p. 201). The "functional" alternative to scientistic evaluation responds to a general, and one would think uncontroversial, definition of rationality[12]: "Maximization

of utility when the labor and costs of calculations and thinking are taken into account" (Good, 1983, p. 23). One simply does what makes sense, or "what works". As Scriven puts it: "The [evaluator's] task is...not to use some experimental design but the best one that is feasible--and that only if it is good enough to establish a worthwhile conclusion" (1983, pp. 76-77).

An Illustrative Example

A number of issues have been considered to this point at a relatively abstract level. Although the next chapter will consider three examples in some detail, a concrete example at this point will help tie the discussions of the quantitative-qualitative and fact-value distinctions together and suggest how educational evaluation can bridge the gap between quantitative and qualitative methods on the one hand and facts and values on the other.

Example. An effort is undertaken to investigate the effects of a medical ethics course required of second-year medical students. The format of the course is 1 hour of lecture per week combined with 2 hours per week of small group discussion. Students meet jointly for the lectures and divide into 6 groups of approximately 10 students each for discussion. Each discussion group is led by a distinct set of preceptors consisting of one physician and one Ph. D. from the humanities or behavioral sciences. Assume that data on learning is collected in 3 ways: direct observations of the discussion groups, interviews with the preceptors and students, and pre-post cognitive testing. Assume the following set of results: (1) Direct observations of the small group discussions and testimony from interviews support the hypotheses that students improved in Knowledge, Reasoning, and the Interpersonal Skills associated with collegial give-and-take. (2) Preceptors, students, and an independent expert question the validity of the cognitive test. A student informant reports that many students, knowing the posttest did not count toward their grade in the course, did not take it seriously and

some intentionally did poorly. (3) An analysis of the pre-post testing in terms of a paired t-test yields a statistically significant negative change.

This study combined quantitative and qualitative methods in two ways. First, quantitative methods (pre-post testing) and qualitative methods (direct observation and interviews) were combined disjunctively to investigate distinct issues: pre-post testing was keyed to cognitive learning, and observation and interviews were keyed to interpersonal skills. Second, quantitative and qualitative methods were combined conjunctively (i.e., as multiple indicators) to investigate the same issue: pre-post testing, observations, and interviews were all used to investigate cognitive learning.

When quantitative and qualitative methods are combined disjunctively, the interpretation of results is relatively straightforward--one simply draws distinct conclusions based on distinct evidence. The more interesting case is when quantitative and qualitative methods are used conjunctively. The conjunctive combination of methods is typically the most puzzling and is the combination that leads thinkers (e.g., Smith, 1983a and 1983b) to question whether quantitative and qualitative methods can be coherently combined.

To reiterate, the contention that quantitative and qualitative are incompatible is an upshot of accepting the positivistic notion that scientific inference consists in building quantitative laws in a mechanistic fashion. Consider the positivistic portrayal of inference in light of

the reasoning that is warranted on the basis of the evidence described in the medical ethics example.

If quantitative methods have the capacity to yield straightforward mechanical conclusions, then the interpretation in the example is clear: the course had a negative impact on students, and the fact that this conclusion is extremely difficult to accept (that it is H-fallible) should be irrelevant to the inference. But quantitative methods have no such capacity precisely because other evidence based on considered (qualitative but L-fallible) judgment may overrule quantitative findings. Contrary to positivism, it is not possible to deduce isolated observation consequences from hypotheses or theories. Not even the results of experiments in the physical sciences have the capacity to coerce conclusions because individual beliefs or hypotheses are never tested directly. An anomalous finding surely requires some adjustment in the belief system, but just where is not something that follows mechanically from quantitative results.

The data from the medical ethics course suggests one especially attractive alternative to the negative impact interpretation, namely, that the cognitive test, the testing procedure, or both, were invalid. The evidence from the observations and interviews supports this interpretation, as does the testimony that the test lacked validity from students, preceptors, and an outside expert. One obviously

does not have good grounds to conclude the course caused cognitive learning. Because of the qualitative findings, however, a decision to dismiss the quantitative findings and thus the hypothesis that the course had a negative impact is warranted.

This conclusion is straightforward and one I trust virtually any researcher would draw, but its obviousness should not obscure two important points. First, the medical ethics study employed quantitative methods, but nothing remotely resembling quantitative law building was involved. Second, qualitative evidence was relevant to the same issue as quantitative evidence, namely, cognitive learning. The two kinds of evidence checked one another, reducing the confidence that could be placed in either alone.

The example shows that quantitative and qualitative methods may be combined both disjunctively and conjunctively. Without falling back on dogmatic methodological assertions, it is difficult to see how either kind of combination of methods is epistemologically suspect.

The combination of quantitative and qualitative methods comprised the descriptive element of the medical ethics study, and this element was inextricably linked to the evaluational element. The medical course was studied in terms of cognitive learning and interpersonal skills. These concepts were used to describe the effects and events of the course (the facts), but they also embodied value judgments. It would have been redundant in the context of the

study to say, "Students learned a great deal _and_ this is good." 'Cognitive learning' and 'interpersonal skills' were not just convenient descriptors, but valued and defensible ends of the course, worth gathering the facts about.

The use of quantitative data in the study does not undermine these value commitments or the ability to make claims about whether desired ends are achieved. Whether a concept is quantified is not directly linked to whether it is evaluative. (Consider judging a diving contest or assigning grades to students.) The data on cognitive learning from the medical ethics study does not mysteriously shift back and forth between being value-free and value-laden, depending on whether the quantitative or qualitative data are at issue.

In summary, the medical ethics course evaluation was driven by judgments about what educational ends are worthwhile. Quantitative and qualitative methods were combined to determine whether these ends were being achieved. There is (or should) be little difference between the general principles that guided this study and those that guide educational evaluation more generally. Again, the question, "How can you _evaluate_ ethics?" may be answered, "The same way you evaluate anything else."

The Presumed Superiority of the Criteria
and Measures of Moral Psychology

The third general feature of scientistic evaluation is a preoccupation with being scientific that leads scientistic evaluators to borrow the measures and mimic the methods of the social sciences, especially psychology. Research on moral education is no exception. The previous section on the quantitative-qualitative distinction addressed research methods; this section will bracket that issue and focus on general approaches to measuring the characteristics of interest. This marks a partial return to the specification of goals and evaluative criteria--the question is whether moral psychology can and should provide the evaluation of medical ethics with such criteria.

Moral psychology is customarily divided into three schools: psychoanalytic, behaviorist, and cognitive moral development (e.g., Hall and Davis 1975; Wren, 1982). Psychoanalytic theory is rarely considered in connection with moral education because its relationship to education as normally conceived is tenuous and because instructors can not be expected to be (nor should they be) therapists. Accordingly, the focus will be on behaviorism (including its near cousin, social learning theory) and Kohlberg's cognitive development theory. The aim is to show that the strategy of borrowing from these two schools of moral psychology is ill-advised in the present context; that whatever their merit as psychological theories, they have

little to contribute to the evaluation of medical ethics
teaching.

Behaviorism

Behaviorism is directly linked to an issue previously
discussed. The inappropriateness of behaviorist-inspired
evaluative criteria was mentioned in connection with
Callahan's criticisms of moral behavior as a goal of medical
ethics teaching. It was observed that Callahan's rejection
of the criterion of moral behavior is well-taken provided
that 'moral behavior' is construed in the behaviorist's
sense. Further support for that observation is provided
here.

Crudely put, behaviorism was the attempt to render
psychology scientific by eliminating reference to
non-observables like intentions. Strongly influenced by
positivism, behaviorists sought to mimic the methodology of
physics (which ironically the positivists had gotten all
wrong). Not surprisingly, the behaviorists' project failed,
both practically and theoretically. On the one hand, the
attempt at "thin" description, as it were, proved
impracticable (Mackenzie, 1977). On the other hand,
behaviorism's positivistic methodological constraint on what
language was to count as permissible involved a fundamental
philosophical flaw.

The untenability of the observation-theory distinction
was especially striking in social science. The elimination
of all reference to the unobservable and mental--the

elimination of things such as reasons, motives and volitions--renders it impossible to distinguish between what Melden (1966) terms "actions" on the one hand and mere "bodily movements" on the other. To use Melden's own example, the elimination of intentions makes it impossible to distinguish a person's arm simply rising (by reflex, for instance) and a person purposively raising their arm (to signal a turn, for instance). Especially relevant to moral behavior, it also becomes impossible to distinguish behavior on the basis of different intentions within the class of "actions". For example, a physician may tell a patient the truth either to avoid malpractice suits or to show respect for the patient, depending on whether the motive is self-regard or other-regard. At the level of overt behavior, (i.e., the movement of telling patients the truth) which motive is involved is beside the point. And this is the problem for behaviorism, because there is a clear and fundamental moral difference between acting out of regard for self and acting out of regard for others--a difference that 'overt behavior' cannot possibly capture.

Social learning theorists who eschew intentions in favor of empirically based "theoretical constructs" do not provide an advance over behaviorists. Consider how Rushton defines the moral concept 'altruism': "Social behavior carried out to achieve positive outcomes for another rather than for the self" (1982a, p. 429). This definition is supposed to exclude intentions, to be stated in "objective, behavioral

terms", but as Krebs points out, in order to avoid reference to intentions, the following kind of definition is required: "Social behavior that achieves positive outcomes, that is, any behavior that produces a positive consequence" (1982, p. 449). The problem with such a definition is manifest:

> It is difficult to imagine anyone arguing that a person who intended to kill another person but in the process shot a malignant tumor out of the victim's stomach (and therefore, in effect, helped the person) was behaving altruistically" (Krebs, 1982, p. 449).

Rushton avoids this absurd result only by doing what he claims not to be doing, namely, making reference to intentions--the expression "carried out to achieve positive results" smuggles them in.

It is ironic that some of the most telling criticisms of behavioristic research methods are to be found in the works of so-called behavioristic philosophers. Gilbert Ryle, usually classified as a behaviorist, contends that behavioristic psychology takes one of two forms: mechanical and para-mechanical (1949). The mechanical form discards the mental altogether (e.g., radical behaviorism); the para-mechanical form allows mental concepts but requires that they be inferred from observation of overt behavior (e.g., social learning theory). According to Ryle, both the mechanical and the para-mechanical forms are based on an illicit bifurcation between overt behavior and the intellect. On the one hand, identifying a piece of behavior as intelligent requires more than a description of mere

movements; some intention must be attributed to the agent as well as the disposition to behave similarly in similar circumstances. This eliminates the mechanical model. On the other hand, intentions and dispositions are not inferred on the basis of <u>independently described</u> overt behavior. That is, the description of behavior itself requires attributing intentions and dispositions (if the behavior is intelligent) or withholding such attributions (if the behavior is a movement). This eliminates the para-mechanical model. As Taylor (1964) observes, behaviorists and near behaviorists face a dilemma of either being saddled with an intentionless language that is too descriptively impoverished to capture anything interesting about human behavior or a language that surrepticiously incorporates intentions.

One wonders why the attempt to avoid intentions is so persistent in light of these difficulties, and there appear to be two primary reasons. Avoiding intentions is believed to contribute to the goal of rendering psychology scientific by (1) reducing fallibility (i.e., increasing objectivity) and (2) eliminating unsupported theoretical constructs (e.g., Rushton, 1982b). Neither of these reasons is convincing in light of criticisms of positivism.

<u>Fallibility of Attributing Intentions</u>. The belief that the attribution of intentions is subjective and unscientific is on a par with (indeed it is an instance of) the criticisms advanced against qualitative data. Drawing on

the earlier discussion of the quantitative-qualitative distinction, it can be dismissed for the same kinds of reasons.

Attributing intentions often depends on testimony, and much is made of this. Rushton (1982b) gives the example of a defendant's testimony in a criminal trial as evidence for the tenuousness of testimony about intentions. Ironically, the doubt in such cases about one intention depends on confidence about another. In the case of the criminal defendant, doubt about the veracity of testimony is based on confidence in the intention to avoid punishment. (Here one is reminded of how unconvincing John Ehrlichman's claim was that only he could know his intentions concerning the Watergate scandal.) Intentions have explanatory and predictive force. If, for instance, it turns out that a piece of overt behavior is self defense and not murder, then, aside from the moral difference in these two actions, one can predict that releasing the accused will not lead to harm to innocent members of society. Likewise, knowing that physicians tell patients the truth out of respect for them as persons (rather than to avoid lawsuits) is a basis for predicting that patients will be treated well in general.

The notion that attributions of intentions are inherently H-fallible is a bit of positivistic dogma, not an obvious truth. Examples drawn from the criminal law are extreme and are by no means paradigmatic. There is often little reason to doubt the inference from agents' testimony

to their intentions. As Wilson observes,

> I can be certain that Churchill did not intend to give into
> Hitler and I can be certain that the reason why my wife went to
> town yesterday was to buy a hat—not because of any scientific
> procedure, but (briefly) because they said so and there is no
> reason to suppose them insincere; moreover, their behaviour
> gives me supporting reasons for what their intentions were
> (1967, p. 210).

Even granting that attributions of intentions are less

reliable than identifications of overt behavior, it does not

follow that intentions should be eschewed—this would amount

to succumbing to the "tyranny of reliability" (Messick,

1981). Increased reliability alone does not entail reduced

fallibility regarding the question of interest. Suppose the

interest were in the percentage of physicians who tell

patients the alternatives to radical mastectomy in the

treatment of breast cancer in order to determine physicians'

respect for patients' rights. Although it is altogether

possible to get a reliable measure of the degree to which

physicians communicate treatment alternatives, the

possibility of systematic error is real. Without knowledge

of why physicians tell patients about treatment

alternatives, inferences about whether physicians respect

the rights of patients (even assuming the measurements are

reliable) are precluded. Some might indeed have respect for

the rights of patients, but others could be trying to avoid

lawsuits, trying to please colleagues, or simply trying to

keep their patients happy. Investigating only the overt

action of communicating alternatives overlooks these

distinctions, reducing the accuracy of generalizations to

other situations where the intentions would be pivotal.

   Intentions as Theoretical Constructs. Behavioristic
theorists remain wedded to a positivistic construal of the
distinction between observation and theory: intentions are
admissible only if they may be reduced to observation.
Rushton (1982b, p. 461) claims that using intentions begs
the question by requiring "definitions of behavioral
phenomena to incorporate a favored explanatory concept"; he
contends:

> Intentions, like motivations, or needs, or stages of
> development, or attitudes, or beliefs, or moral principles, or
> any other intangible hypothetical construct, are themselves
> inferred from regularities in behavior (1982b, p. 460).

The view exemplified presupposes that intentions can be
clearly distinguished from and set over against descriptions
of actions in the same way the positivists believed that
theoretical constructs could be set over and against
observations more generally. This view may be criticized in
two ways.

   First, it presupposes the untenable positivistic
distinction between theory and observation. Observations
(i.e., descriptions) of behavior are, in a manner of
speaking, "intention-laden". For example, 'murder',
'voluntary manslaughter', 'involuntary manslaughter', and
'justifiable homicide' all name different actions. To
attribute one intention rather than another is to identify
one action rather than another. The relationship between
intentions and actions is thus not contingent. It therefore
would make no sense to investigate whether the intention to

help others regularly accompanied altruistic actions because
the intention is definitive of the behavior. The charge
that to use intentions in describing actions "begs the
question" is thus incoherent. Inferring intentions from
"regularities of behavior" is also incoherent, since
identifying a regularity presupposes classifying a set of
actions which in turn presupposes attributing intentions.

Second, it is a distortion of major proportions to lump
such things as intentions, moral principles, and the like,
under the rubric "hypothetical constructs"--a rubric which
would include such things as the id, quarks, and
anti-matter. Intentions and moral principles are not
posited in order to explain behavior causally; they have
instead to do with describing and guiding it respectively.
Social scientists are not free to use these (or
operationalize them) in any way they see fit, for such
concepts have entrenched ordinary meanings. Distortion,
irrelevance, and insidious bias are bound to result when
operationalizations trade on, but are not faithful to,
ordinary meanings of action descriptions.

Conclusion. The discussion of behaviorism may now be
brought to bear on the evaluation of medical ethics
teaching. First, the claim that behavioristic approaches
are the only means by which to obtain credible empirical
knowledge about moral behavior is unconvincing. The claim
is rooted in the tenets of positivism and subject to the
standard criticisms advanced against that now moribund

construal of science.

Second, the evaluation of ethics teaching is concerned with the kind of ordinary language meaning attached to descriptions of behavior like Moral Regard, Empathy, Reasoning, and so on, not with theoretical, intentionless concepts. Even if there is something to be gained in psychological theorizing by developing a set of "theoretical constructs"--and, arguably, there is not--such a vocabulary has little usefulness in the context of evaluating ethics teaching.

Third, a primary aim of ethics instruction is the development of "discursive moral competence" (Ruddick, 1981). Ethics instructors seek to enhance the ability of their students to discourse about and provide reasons for their ethical views and behavior. Extending Ryle's position, Wren (1982) argues that because behavioristic approaches take either mechanical or para-mechanical forms, they are incapable of capturing the prescriptive and self-regulating nature of moral behavior. For behavior to have a moral dimension, reference has to be made to some prescribed norm. Praise (blame) follows from regulating (failing to regulate) one's self in terms of the norm. Mechanical and para-mechanical explanations depend only on which desire wins out; they leave completely out of the picture how reasons might serve to regulate behavior and in the process overrule one's strongest desire.

Kohlberg's Theory

Kohlberg's theory of moral development [13] has been pummeled with criticisms,[14] but the criticisms differ from those lodged against behaviorism. Unlike behaviorism, his research seeks to investigate cognitive moral judgment: ethics instruction's primary focus. He explicitly acknowledges the importance of moral philosophy, and, according to Puka, "Cognitive-developmentalism is no less than an attempt to empiricize ideal models of moral rationalism" (1982, p. 471).

Despite the promise of Kohlberg's theory as an alternative to behavioristic moral psychology, there are three compelling reasons for viewing it as ill-suited for evaluating medical ethics teaching: insensitivity to instructional effects, lack of interpretability, and implicit moral hegemony.

Insensitivity to Instructional Effects. Medical ethics instructors stimulate students to be more autonomous and less conventional in their moral thinking. In Kohlberg's terms, the aim is to move students from "level 2" to "level 3", so the theory captures one important aspect of ethics teaching. This aspect, however, is too general to have practical significance. Although pre-post gains in ethics courses have been reported using Kohlbergian measures (Rest, 1979), this is exceptional. Kohlbergian measures typically fail to demonstrate instructional effects (Lickona, 1980 and Rest, 1982). A medical ethics instructor would likely view

getting students to understand informed consent--the medical
ethics literature addressing it and its reasoned
justification--as considerable progress, yet such progress
would be undetectable in terms of Kohlberg's levels and
stages.

The explanation might be that ethics teaching simply
doesn't work, i.e., that there are no effects to be
measured. Although plausible, this conclusion is hasty at
best. Given the formal, i.e., content-independent,
cognitive "structures" Kohlberg's theory posits, it is not
surprising that measurable progress would not follow from
the relatively brief exposure typically afforded by ethics
instruction. A more reasonable conclusion is that
Kohlberg's structures are too abstract to detect progress at
more specific levels.

For instance, in the present version of Kohlberg's
theory (Kohlberg, 1982) the highest stage is the
post-conventional and the next highest is the conventional,
"law and order" stage. Progress would have to indeed be
dramatic to move from the lower to the higher of these
stages. One would also expect much variation within the
post-conventional stage (e.g., two individuals, both at the
post-conventional stage, would disagree about whether cancer
patients should be told their diagnosis if they disagreed
about whether such disclosures are harmful). Furthermore,
it is doubtful that critical reasoning in general is as
formal as Kohlberg's theory takes moral judgment to be

(McPeck, 1980), and empirical studies indicate that medical decision-making (Elstein et al. 1978) and moral reasoning (Iozzi and Paradise-Maul, 1980) are not.

Lack of Interpretability. Even if progress were detected using Kohlbergian measures, it is not clear what to make of it. What, for example, does a one-quarter stage gain on James Rest's DIT mean?[15] If presented with these results, the ethics instructor is likely to press for a clarification--Does that mean that the students now see the reasons for the presumption in favor of telling their patients the truth? The evaluator will be at a loss. Although the problem of test interpretation cannot be altogether eliminated--What, for example, does a 5 point gain on a 50 item test mean?--interpretation is more straightforward when familiar reference points are available. A gain of 5 points on a 50 item test is equal to a gain of 10%, for instance, and may be the difference between a "B" and an "A-". In addition, when tests possess "face validity" (i.e., when the relationship between tests and the objectives of instruction is clear by inspection), interpretability is enhanced. For example, if students do poorly on items having to do with the concept of patient autonomy, then it may be inferred that the teaching of the concept is in need of improvement. Similar uses for Kohlbergian-type tests are unavailable.

Implicit Moral Hegemony. Kohlberg's theory is biased against women (Gilligan, 1977) and against the religious
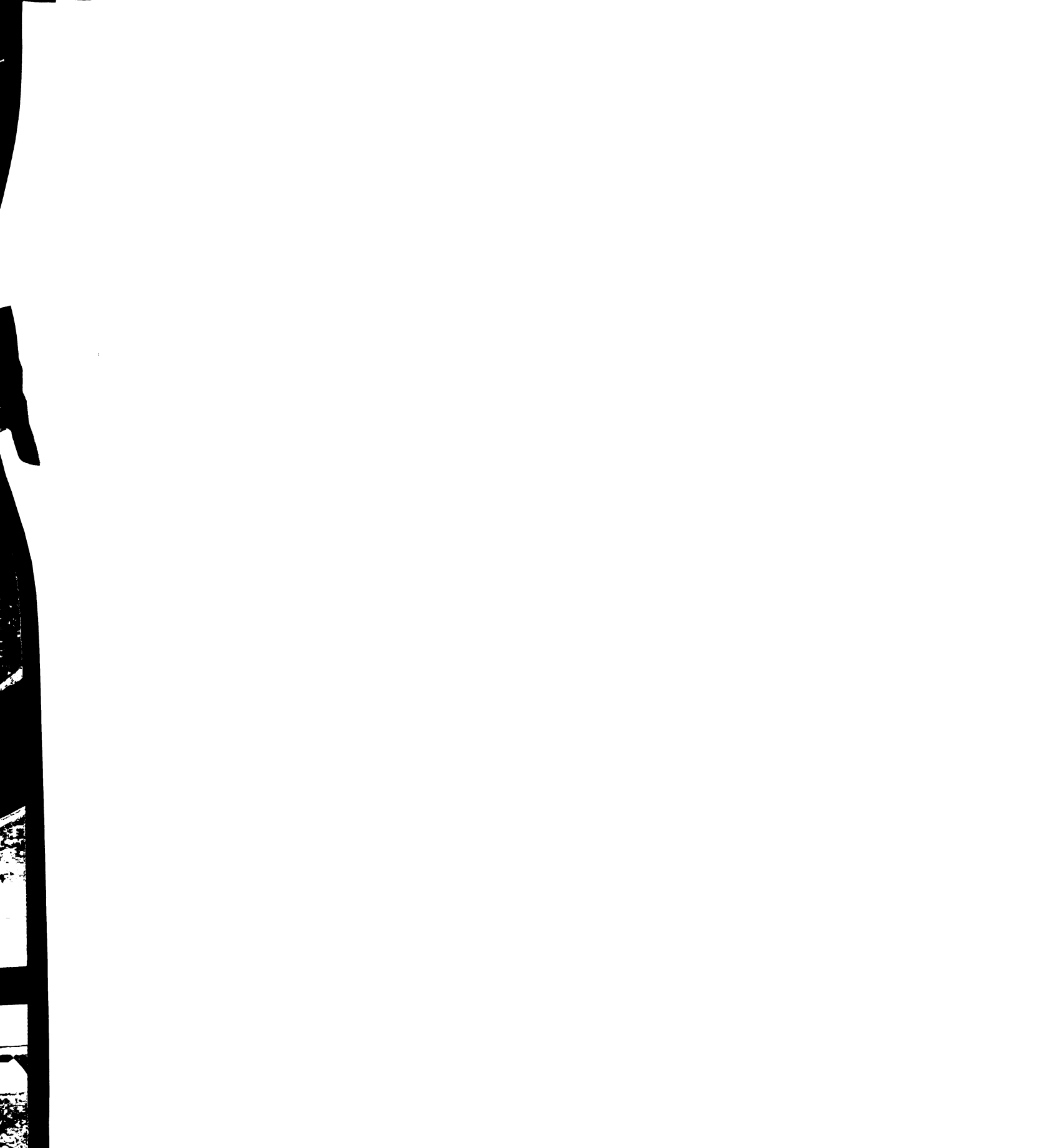
(Dykstra, 1981 and Flanagan, 1982). This by itself raises a serious moral question about the use of Kohlberg's theory as the basis for evaluating students or programs; it is also symptomatic of a more fundamental problem. By imposing a pre-set sequence of stages of development, where "higher" equals better, the stages become authoritative and effectively cut off discussion: judging the quality of a moral position becomes a matter of determining what stage it exemplifies. Although it is perfectly acceptable to describe and classify moral judgments in the research context, such descriptions and classifications cannot be straightforwardly converted into prescriptions for judging either individual moral positions or ethics teaching. As argued in the previous chapter, moral reasoning is evolving and self-critical, and "pre-established blueprints" are ruled out of court.

Conclusion

This chapter endorsed functional evaluation and considered scientistic tenets that imply a functional approach is somehow epistemologically suspect or unscientific. Three obstacles were identified which are especially relevant to evaluating ethics teaching: a rigid fact-value distinction, a rigid quantitative-qualitative distinction, and a presumption in favor of criteria and measures grounded in moral psychology.

A rigid fact-value distinction underlies the demand that social research be descriptive only. If the distinction and demand are legitimate, the consequences for evaluating medical ethics teaching are disastrous. Ethics teaching is pre-eminently concerned with examining and defending moral judgments; a prime objective of such teaching is thus a species of just what is declared illicit--value judgments. But the grounds for a fact-value distinction that renders value judgments non-cognitive and undecidable went by the boards with positivism. There are no good reasons to view value judgments in general and moral judgments in particular as essentially non-cognitive. The demand that social research be wholly descriptive went by the boards as well. The concepts employed in educational research in general--achievement, intelligence, and the like--have evaluative interpretations. There is thus nothing particularly objectionable or epistemologically suspect about using clearly value-laden concepts, such as the Wilson+1 goals, in ethics teaching and its evaluation.

For scientistic evaluators, quantitative methods are identified with science and objectivity, and qualitative methods are identified with non-science and subjectivity. This way of distinguishing the methods presupposes the untenable positivistic distinction between theory and observation. There is no a priori epistemological difference between quantitative and qualitative methods that warrants judging one or the other uniformly superior.

Reducing the fallibility of beliefs about issues deemed important is the only defensible criterion for assessing methods. Given this criterion, whether quantitative methods, qualitative methods, or some combination is best (rational, scientific, and so on) cannot be answered in the abstract; it depends on the nature of an investigation and the constraints under which it must be conducted. The evaluation of ethics teaching thus cannot be criticized a priori for employing or failing to employ this or that type of research method.

Finally, the demand to use instruments based on moral psychology is not defensible. The behavioristic criterion of overt behavior is fundamentally flawed, inimical to the teaching of ethics, and subject to the same general criticisms that apply to positivism. Kohlberg's cognitive developmental theory is beset with practical problems when proposed as a basis for evaluating medical ethics teaching.

Because formal instruction in medical ethics is relatively new and involves complex and poorly understood aims, a flexible approach to evaluation is indicated—an approach which begins with the nature of medical ethics teaching and works its way outward. Chapter II explicated the nature of medical ethics teaching; this chapter has overcome obstacles to working outward from that nature. The general criterion of the merit of evaluation is reducing fallibility about questions of interest. There are no good reasons to avoid value judgments, qualitative methods, or

non-scientific concepts in the evaluation of ethics teaching

(or any other kind). On the contrary, each of these is

indispensible.

NOTES

1. "Scientistic" is borrowed from Cronbach (1982). He is not committed to my use of the term.

2. It is puzzling that many who would endorse this claim go on to derive the position, considered toward the end of Chapter II, that we ought not be teaching ethics at all. They seem oblivious to the fact that they are precluded from making any ought-statements whatsoever. Their problem is similar to the determinist's who, after arguing that all actions are determined and therefore beyond our control, advises us not to cry over spilt milk.

3. For those who are unfamiliar with the term, 'reliability' roughly means consistency. For example, if a test renders dramatically different scores for an individual on two administrations within a reasonably short period of time, its test-retest reliability is low; if two individuals assign markedly different scores to the same test, its inter-rater reliability is low (a common problem with essay exams). The term 'validity' roughly means fidelity. It is commonly held that some degree of demonstrated reliability is a necessary condition of validity.

4. Lipman et al. (1977) criticize "values-clarification" for having this characteristic. The hidden message is that values have no cognitive, rational dimension amenable to critical scrutiny.

5. Mercer (1971), for instance, documents the effects of intelligence testing on educational placement of groups such as Hispanics and blacks.

6. Refusal by parents to consent to surgery for Down's infants with correctable life-threatening anomalies is based in large measure on the expected intellectual functioning of such infants. Whether such refusals are morally justified has been a topic in the medical ethics literature for at least a decade (e.g., Shaw, 1977 and Rachels, 1975).

7. The distinction between qualitative and quantitative is by no means clear, if for no other reason than the sheer number of descriptors used to characterize it. For example, below is a partial list adapted from Reichardt and Cook (1979, p. 10).

QUALITATIVE:	naturalistic and uncontrolled observation
	subjective
	close to the data; the "insider perspective"
	grounded, discovery-oriented, exploratory, expansionist
	descriptive, and inductive
	process-oriented
	valid; "real", "rich", and "deep" data
QUANTITATIVE:	obtrusive and controlled measurement
	objective
	removed from the data; the "outsider perspective"
	ungrounded, verificationist-oriented, confirmatory, reductionist, inferential, and hypothetico-deductive
	outcome-oriented
	reliable, "hard", and replicable data

8.   The "analytic-synthetic" distinction was crucial to logical positivism's concepts of "cognitive significance" and "verifiability". Analytic statements were defined as ones whose truth-value could be determined solely by appeal to logic and meanings (e.g., All bachelors are unmarried). Synthetic statements were defined as ones whose truth-value could be determined solely by appeal to observation (e.g., Grass is green). Any statement which was to count as "cognitively significant" (i.e., legitimate in science) had to be verifiable either in terms of observation (if it were synthetic) or in terms of logic and meanings (if it were analytic).   All other statements were barred. The class of illegitimate statements included "metaphysical" ones, e.g., 'God exists', and "emotive" ones, e.g., 'Abortion is morally wrong'.

Quine (1962) attacked the analytic-synthetic distinction directly and convincingly. In the process, he paved the way for the kind of post-positivistic interpretation of science made prominent by Kuhn in which all scientific knowledge is seen to be theory-laden and not neatly divisible into the purely observational (or synthetic) and the purely theoretical (or analytic).

9. The much lamented gap between theory and practice is partially attributable to the vocabularies social researchers choose. Rorty (1982b) sets down two requirements for the vocabulary of social science:

> (1) It should contain descriptions of situations which facilitate their prediction and control.

> (2) It should contain descriptions which help one decide what to do (p. 197).

Service to the first requirement in the name of science detracts from attention to the second, and this reverses priorities. Given the aim is to improve practice, the first order of business is to ensure that concepts employed are useful toward this end. That such concepts might be mundane, <u>unscientific</u>, and value-laden is an objection only if the question of purposes is begged.

10. Terms such as 'knowledge' and 'belief' are "opaque". Unlike "transparent" terms, e.g., 'is greater than', opaque terms are not "well-behaved" in formal logical and mathematical languages. The use of opaque terms requires judgments of relevance, "similarity judgments", whereas transparent terms do not (Quine, 1969b). Related to this, Toulmin (1960) argues that physicists are free to invent vocabularies and a well-behaved mathematical system to suit their aim of investigating "the form of given regularities". By contrast, "natural historians" (which presumably includes social scientists) are saddled with a less technical and public vocabulary fitted to their aim of investigating "regularities of given forms" (p. 53).

11. Notably, this is consistent with Kuhn's remarks about social research (1961). Despite the manner in which his notion of a "paradigm" has caught fire among educational researchers and been used to contrast quantitative and qualitative methods, he excluded social science (let alone individual methods) on the grounds that it has never had an accepted core of theory required to constitute a "scientific paradigm" or to make for "scientific revolutions".

12. More specifically, Good calls this "Type II Rationality" or the "Principle of Non-Dogmatism". I don't claim that the principle is easily applied.

13. Kohlberg's theory construes moral development in terms of progression through successively more adequate "levels" (with two "stages" each) of cognitive moral judgment. The progression through levels and stages is held to exist across cultures. Although few individuals reach the highest stage, progress is putatively irreversible. See Lickona (1980, p. 106) for a description of the stages and levels in Kohlberg's theory. Also see Kohlberg (1982) for the latest

version of "level 3" which now has only one stage.

14. Lickona enumerates the following litany of criticisms, all of which he claims have merit:

> [Kohlberg] has been taken to task for centering too much on justice in defining the moral; for being culturally biased in his definition of morality; for being sex-biased in studying an exclusively male sample and for defining the stages in ways that emphasize "masculine" themes of rights and justice and neglect "feminine" themes of responsibility and love; for going from a description of what moral development is to a prescription of what it ought to be; for devaluing conventional morality; for overestimating the role of reasoning in moral functioning, and underestimating the role of other factors, such as affect, personality, habit, and expectations of consequences; for failing to take into account adequately the impact of the particular nature of the moral dilemma on the stage of moral reasoning elicited from a subject; for lack of sufficient validity and reliability of his research methodology; for having insufficient evidence for his two highest stages; and for failing to respond to his critics (1980, pp. 107-08).

15. The DIT (Defining Issues Test) is a paper-pencil version of several (6) of the moral dilemmas used by Kohlberg in his research. It was developed by one of Kohlberg's students, James Rest. See Rest (1979) for a description and thorough review of the research.

CHAPTER IV

EVALUATING MEDICAL ETHICS TEACHING

The general goal of educational evaluation is reducing fallibility about questions of interest. Chapter II specified the questions of interest peculiar to medical ethics teaching, and Chapter III suggested considerations that go into reducing fallibility. This chapter employs three concrete examples of course evaluations to demonstrate how the abstract considerations of these earlier chapters may shape the actual practice of evaluation. The chapter aims to show that valid and credible evaluation of medical ethics teaching is possible and can serve both of the customary functions of evaluation--improving practice and measuring achievement of goals.

The chapter is divided into two major sections. The first specifies a "basic functional model" that is keyed to the special nature of courses (versus other modes of instruction). The model is grounded in the Wilson+1 goals and combines qualitative with quantitative research methods. The second section uses variants of the model to frame discussions of three course evaluations: a nursing ethics course, and two successive offerings of a medical ethics course.

111

A Basic Functional Model for the
Evaluation of Medical Ethics Courses

Goals Revisited

In Chapter II the goals of medical ethics teaching were developed by means of a four-step procedure: the ends of medical ethics teaching were described in the abstract, modified to accord with the typical constraints on educational practice, compared to expert opinion, and then assessed in terms of "controversies" and "misconceptions". Further refinement regarding practical constraints is required when courses are the particular object of evaluation.

Given the typical context for courses—one in which students meet in groups for discussions or lectures and do not encounter patients—the direct goals, Knowledge, Reasoning, and Appreciation, are the focus; the indirect goals, Moral Regard, Empathy, Interpersonal Skills, and Courage, fade into the background. There are several reasons for this emphasis.

Within courses, instructors (or evaluators) are unable to make very careful observations of the behavior of individual students and are altogether precluded from determining how students respond to patients. Accordingly, the indirect goals apply only in so far as they relate to the issue of how effectively and sincerely students engage in classroom activities. Although the indirect goals remain important in this special sense (since collegial discussion,

for example, is an important means by which ethical problems are addressed in the practical setting of hospital ethics conferences), the aim of enhancing students' ability to interact with and respond to patients in morally appropriate ways is, as a practical matter, secondary.

On the other hand, the practical disadvantages of courses regarding indirect goals are counter balanced by advantages regarding direct goals. Courses provide an especially suitable context to emphasize the cognitive nature of medical ethics and instill an appreciation for it. They provide an economical and effective means to provide the necessary cognitive foundation that may be built upon and reinforced in other, less controlled contexts, which involve numerous other educational aims.[1]

In light of these considerations, the model and the examples of this chapter will heavily emphasize the direct goals. Only minor attention will be paid to the indirect goals and only in one of the three examples.

Methods

The general methodological problems are to (1) fit data collection techniques to the relevant goals and (2) develop research designs that have a reasonable chance of sufficiently reducing fallibility. Four means of data collection (tests, direct observation, student evaluation forms, and interviews), coupled with a "multiple indicators" design strategy (explained later), make up a suitable model

for most contexts.

Data Collection

1 _Tests_. It is customary to measure cognitive achievement with paper-pencil tests. Although the practice is not without problems (test wiseness, bias, teaching to the test, and invalidity, for instance, defeat the purposes of testing), using tests is often indicated both on intuitive grounds and because suitable alternatives do not exist. Testing in ethics is not fundamentally different from testing in anything else. It is an appropriate means of evaluating the cognitive aims of medical ethics teaching provided it avoids the customary problems and is appropriately keyed to the content and skills of interest (i.e., Knowledge and Reasoning respectively).

2 _Direct Observation_. Unstructured direct observation (structured observation is distinguished below) is a second method of data collection, and it, like tests, may also be used to trace progress relative to the goals of Knowledge and Reasoning (but in the context of discussions of ethical issues as opposed to performance on examinations). Unstructured direct observation is also useful for examining Appreciation and the indirect goals (such as collegial discussion as a variant of Interpersonal Skills).

The primary advantage of direct observation is its capacity to detect what is unanticipated and therefore not amenable to pre-designed measures; its primary disadvantage is that it permits the idiosyncracies of observers to

determine the focus of data collection and to influence interpretation.

In general, the more unstructured observation is, the more economical, unrestrictive, and limited in generalizability it will be; the more structured observation is, the more expensive, restrictive, and generalizable it will be. Trade-offs must be made which balance these considerations against the nature of the question and its fallibility. Structured observation (e.g., checklists of overt behavior) is indicated where there is a desire to narrow the focus in order to perform comparisons across large groups; unstructured observation (e.g., simply observing activities and taking notes) is indicated where the focus is broad or undetermined, and where the scope of the study is small. Relatively unstructured observation is usually indicated when medical ethics courses are the object of evaluation.

3 <u>Student Evaluation Forms</u>. A third method of data collection is student evaluation forms. Used during or, more typically, at the end of a course, student evaluation forms are well suited for measuring attitudes. Strictly speaking, Appreciation cannot be measured directly by attitudinal measures because it has a cognitive dimension in addition to an attitudinal one. Student evaluation forms, however, provide the most convenient and direct information about Appreciation in practical educational contexts. Students are accustomed to filling them out and they are

almost universally employed in higher education to evaluate teaching.

Although frequently criticized for measuring wants rather than needs, student evaluation forms provide a valuable source of information in light of the special commitment to "internal goods" that moral competence requires. Where such student evaluation forms provide for free responses (e.g., "liked best", "liked least"), the wants versus needs issue can be addressed directly because the reasons given for wanting or not wanting, liking or not liking, inform interpretation. For example, "The course made me think" is a kind of reason that would allow going beyond mere wants to infer needs, "The course was entertaining" would not sanction such an inference.

4 <u>Interviews</u>. A fourth method of data collection is interviews. These may range from highly structured protocols, bordering on questionnaires, to informal chats, perhaps prompted by something observed. Interviews, like direct observation, may be used to measure Knowledge, Reasoning, and attitudes. The same considerations regarding the degree of structure that apply to direct observation apply to interviews as well.

Figure 3 keys the four data collection techniques discussed in this section to the effects they are intended to measure.

	Knowledge	Reasoning	Attitudes
tests	x	x	
observation	x	x	x
evaluation forms			x
interviews	x	x	x

Figure 3. The Relationships Among Data Collection Techniques and
Constructs in the Basic Functional Model

Design

Of the four data collection techniques described, only
tests and student evaluation forms are easily amenable to
quantitative analysis. Thus, the suggested model combines
quantitative and qualitative methods. Given the arguments
of Chapter III, there is nothing suspect or incoherent about
such a combination; the only proviso is that the various
means of data collection work together to provide useful
information that enjoys reasonable support.

This combination of data collection methods makes
implicit use of the strategy of "multiple indicators"--
applying various kinds of data to the same research
question(s). Intuitively, the use of multiple indicators
helps protect against the peculiarities of given instruments
or observers, and it helps strengthen inference (reduce
fallibility) in the same way a second opinion or an
additional laboratory test increases (or reduces) confidence
in a medical diagnosis. The use of multiple indicators is

also supported on technical statistical grounds (e.g.,
Cronbach, 1982).

Figure 3 illustrates a multiple indicators approach.
Notice that both the Reasoning x tests and Reasoning x
observation cells are occupied. The two kinds of data cross
check one another and bolster inferences that may be drawn.
For example, if students exhibit improved Reasoning on the
basis of testing and observation or on the basis of neither,
then the inference is stronger than it would be if only one
kind of data were available. If students show improvement
on one but not the other (e.g., the Chapter III example),
the interpretation would be more unclear but the conclusions
reached would be more trustworthy than they would be if
which only one source of data was available.

In addition to strengthening inferences, the use of
multiple indicators also enhances the flexibility of
evaluation. It makes little sense to persist in a
pre-conceived design if, on the basis of information from
other data sources, it seems likely that the design in
question is headed down a blind alley, or if it is suspected
that persisting for the sake of the design compromises the
education of students.

Heading down a blind alley occurs when an instrument,
once in use, is judged invalid. An example of this kind was
discussed previously. A pre-post design was employed to
investigate cognitive effects in a medical ethics course.
Information derived from alternative data sources, namely,

the testimony of students and faculty discussion leaders indicated that the aims of the testing would likely be thwarted because the test lacked face validity. In this situation, it would have been wise to have seriously considered aborting the testing; this would have freed time and resources to pursue more promising avenues.

The second situation, compromising the education of students, occurs when the instrumentation is acceptable but the treatment (i.e., the course) needs changing in a way that would invalidate the pre-measures. An example of this kind arose in connection with an ethics write-up assignment for third-year medical students doing their clinical rotation in medicine. On the basis of feedback from students, it was evident that the written instructions for the assignment were inadequate and the kinds of changes that could be made to improve them were straightforward. The supervisor of the write-up assignment was reluctant to make the needed changes on the grounds that this would compromise comparisons between the performance of initial cohorts of students who had had no training in ethics in their first two years of medical school and later cohorts who would have had such training. Now, changing the exercise did preclude the desired pre-post comparisons, but the decision to proceed with the changes and compromise the design was justified on the grounds that one's first obligation is to provide students with the best education possible. Because the exercise was questionable, the information that could

have been obtained from the original design would be nearly worthless in any case. Preserving the design probably would have reduced fallibility, but not about a question of interest.

In summary, the methods section described four means of data collection--tests, direct observation, student evaluation forms, and interviews--and how these four methods can be combined in a multiple indicators evaluation approach. A basic functional model for medical ethics courses was described in terms of how the four data collection techniques may be keyed to Knowledge, Reasoning, and attitudes. More specific research methods and designs will be discussed in terms the three examples in the next section.

Three Course Evaluations

This section applies the basic functional model to three course evaluations, with the modest aim of showing that medical ethics teaching can be evaluated in a way that meets the reasonable demands of evaluation researchers and ethics instructors. The discussion of each example is divided into six sections: (1) course description, (2) evaluation purposes, (3) evaluation design, (4) evaluation results, (5) evaluation impact, and (6) conclusion. The "course description" sections are straightforward: they describe the content, pedagogical methods, and settings of the examples. The "evaluation purposes" sections further set the stage of

the examples, and establish one criterion for the success of the evaluations, namely, achieving substantive evaluation purposes. The "evaluation design" and "evaluation results" sections emphasize meta evaluation issues, i.e., issues about the effectiveness of evaluation methods (versus substantive evaluation questions, such as the relative merit of the courses themselves). These two sections are the heart of each example, and are most directly linked to the functional evaluation approach that grows out of Chapters II and III. The "impact" sections report the utilization of results and relate substantive evaluation purposes to evaluation effectiveness. The concluding sections summarize what each example shows.

Before turning to the examples themselves, two clarifications about the use of the basic functional model are in order. First, as stated in Chapter I, the perspectives on goals and research methodology defended in Chapters II and III did not dictate the conduct of the three course evaluations, especially not the first two. Instead, the goals and evaluation methods were developed concomitant with, and largely in response to, the concrete evaluations. The examples have been recast in a somewhat ahistorical way to facilitate exposition on the one hand and linkages to Chapters II and III on the other. The "evaluation design" sections in particular incorporate post hoc data analysis decisions that constituted reasonable uses of the data collected, but that were not always envisaged when the

evaluations were in fact designed.

Second, the discussion of the examples does not aim to provide a blueprint for the evaluation of medical ethics teaching or to establish that some alternative (perhaps superior) ways of evaluating medical ethics teaching are out of the question. The basic functional model presupposes that various practical constraints are operative (e.g., that resources are not available to cross-check direct observation among multiple observers, to analyze verbatim transcripts, to record and analyze video tapes, and so forth). The discussions of the three course evaluations are designed to illustrate only how more specific limitations and contextual features than those already contained in the basic functional model can be accommodated by adapting the model.

Example 1: Ethics in Nursing (1980)

Course Description

"Ethics in Nursing" is an elective course offered to junior and senior nursing students by Michigan State's Philosophy Department. It is a two-credit course and spans a ten-week term, meeting once a week for two hours. Development of the course was funded by a National Endowment for the Humanities Grant[2]. The evaluation described here was conducted in 1980, the second time the course was offered.

The written material for the course was Ethics in

Nursing (Benjamin and Curtis, 1981). Most of the instruction focused on Socratic discussions of cases from the textbook; mini-lectures were used when appropriate. The course was team taught by a philosopher and a nurse. Students were evaluated on the basis of their performance on short essay responses to cases and on their contributions to class discussions.

Evaluation Purposes

The evaluation had two primary motivations. First, an evaluation was required by the funding agent. Although the university provided such services upon request, one of the instructors had had previous experience with evaluators provided by the university and found their methods simplistic, trivializing, and atomistic (i.e., he viewed expert evaluation with the same general skepticism that is pervasive among ethics instructors, Caplan, 1980). In the instructor's estimation, the available evaluation experts had virtually no understanding of the nature and aims of ethics teaching. Thus, the instructors welcomed aid from a former philosophy graduate student, re-tooling in evaluation research. Second, the instructors desired to demonstrate the evaluability of this course (and similar ones). Despite their skepticism about the methods typically used in the evaluation of ethics teaching, the instructors earnestly believed that measurable effects, consistent with the course objectives, would result from the course; and they were willing to put this belief to an empirical test. They

simply required that the evaluation fit the nature and aims of the course.

Related to the instructors' desire to demonstrate the evaluability of their course, the design of the Ethics in Nursing evaluation (described in the next section) was motivated by three meta evaluation questions: (1) Can measures of Knowledge, Reasoning, and Appreciation be developed? (2) Can the basic functional model for medical ethics courses (depicted in Figure 3) yield warranted conclusions? and (3) Can the empirical results obtained by employing the model affect the practice of medical ethics teaching?

These three questions are related to the remainder of the Ethics in Nursing example in the following way. At an epistemological level, sensitivity of the design to instructional effects (in terms of the desired goals) is relevant both to whether Knowledge, Reasoning, and Appreciation can be measured and to the value of the basic functional model. At the practical level, the impact of evaluation findings is relevant to the usefulness of the model as a means of gathering information for criticizing and improving medical ethics teaching.

Evaluation Design

Data Collection. Data for the Ethics in Nursing course were collected by the four means listed in the model of Figure 3, namely, observation, student evaluation forms, interviews, and tests. The course was observed eight of the

ten times it met, usually in the second of the two hours
making up each meeting. Two student evaluation forms were
used: the University's "Student Instructor Rating Form"
(SIR), a two-part instrument with fixed-response items and a
comments section, and a mailed fixed-response instrument
designed especially for the course. Informal interviews of
the instructors were periodically conducted. Finally,
students were required to write an ungraded essay in the
first and last meetings (The essay was an analysis of
Margaret's dilemma, the example of Chapter II).

Interviews of the instructors did not yield information
pertinent to the present discussion. Figure 4 depicts the
relationships among the constructs and the other three
methods of data collection.

	Knowledge	Reasoning	Attitudes
essay tests	x	x	
observation	x	x	x
student evaluation forms			
SIR			x
mailed			x

Figure 4. Relationships Among Data Collection Techniques and
Constructs For Ethics in Nursing Evaluation

The Design and its Logic. For cognitive effects, the
basic research design was a case-study, Observation-
Treatment-Observation (O X O). Resources were not available
to arrange a comparison group, and even if there had been,
the self-selected nature of the group who elected the Ethics
in Nursing course would render a comparison group suspect.

Moreover, the course was sufficiently distinctive in the curriculum to make competing explanations for observed results implausible.

For student attitudes, the basic research design was contingent on the means of data collection: Time-Series (O X O X O) for direct observation and After-Only (X O) for student evaluation forms. Reasoning similar to that used to justify the design for cognitive effects justified foregoing control groups in investigating students' attitudes--given the nature and setting of the course, competing explanations for observed effects would be implausible. Feasibility (primarily resource constraints) was also a consideration in the designs.

These individual designs combined with the three kinds of data (testing, observation, and student evaluation forms) to form the following logical framework for interpreting the data.

Data from essay testing and direct observation focused on measuring Knowledge and Reasoning, and created the foundation for a three-step analysis. (1) Statistical comparison of the pre-post essays would provide evidence for or against cognitive effects in terms of Knowledge and Reasoning and, at the same time, evidence for or against essay tests as a means of measuring Knowledge and Reasoning. (2) Depending on the results of the statistical analysis, close inspection of the essay tests in terms of specific criteria associated with Knowledge and Reasoning would lend
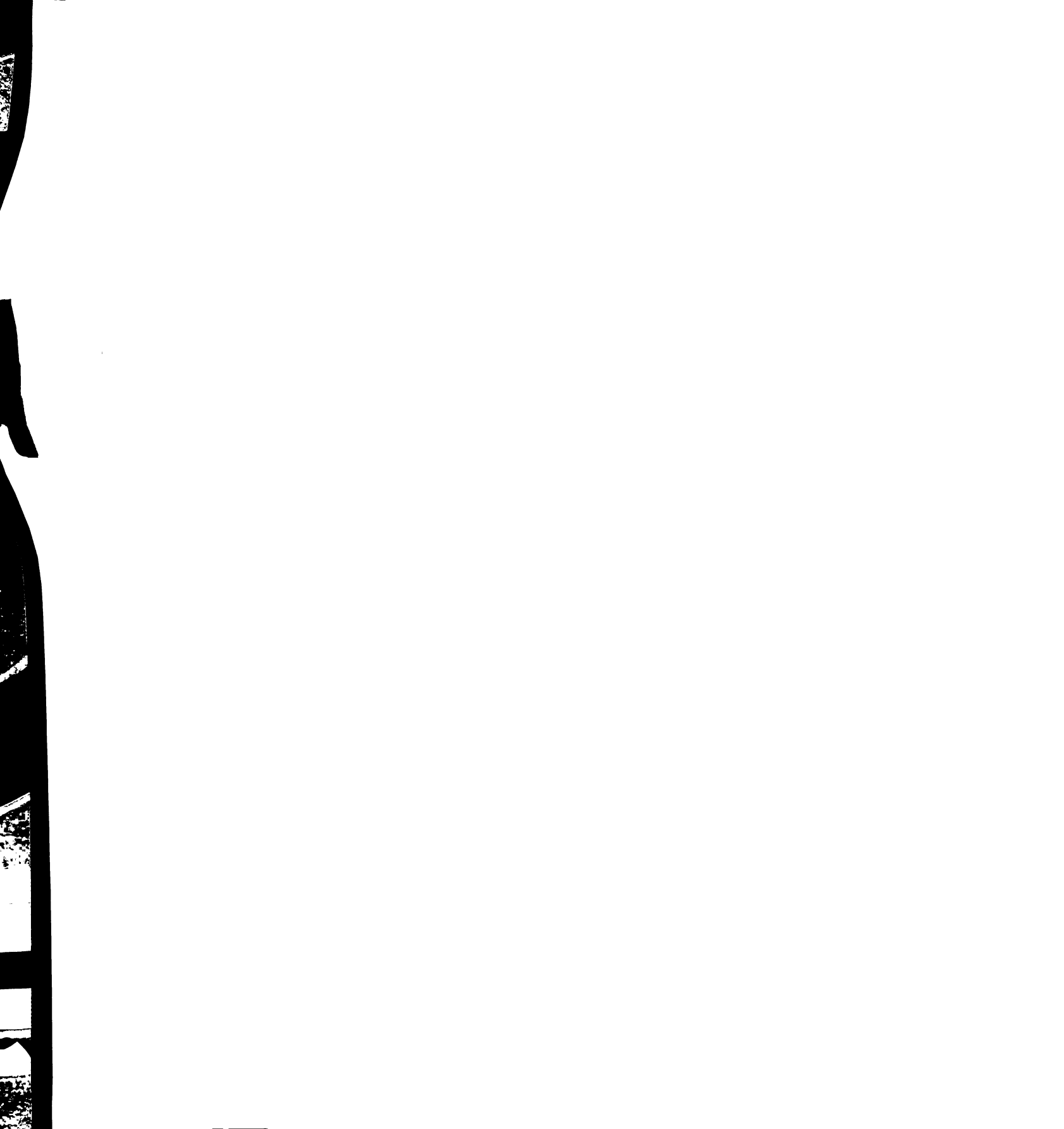
interpretability to the statistical analysis. If, for example, the statistical test was significant in the desired direction, documenting that the difference was explicable in terms of criteria associated with Knowledge and Reasoning would help ensure that the essays were valid measures. If the statistical test was not significant, close inspection of the essays might reveal that the statistical test was insensitive to effects that can be documented by alternative means. (3) Depending on the results of the two kinds of analysis of the essay testing, pre-post comparisons based on direct observations of student performance in discussions (also keyed to Knowledge and Reasoning) would support or count against the findings from testing.

Direct observations of discussions and student evaluation forms were used to measure students' attitudes. As explained previously, these data collection techniques cannot directly measure the real target, Appreciation, because Appreciation includes both cognitive and attitudinal components. Thus, positive attitudes alone (not accompanied by evidence for cognitive effects) cannot be distinguished from mere wants. Nonetheless, because the logic of the Ethics in Nursing design incorporated independent means of investigating cognitive effects, accommodating Appreciation within the overall design was largely a matter of adding the measurement of student attitudes to the measurement of cognitive effects.

<u>Evaluation Results</u>

As stated previously, this section is keyed to the related questions of whether Knowledge, Reasoning, and Appreciation can be successfully measured and whether the basic functional model can yield warranted conclusions. These questions will be addressed first with respect to Knowledge and Reasoning and then with respect to Appreciation.

<u>Knowledge and Reasoning</u>. Primarily two kinds of data were used to explore Knowledge and Reasoning: essays and direct observations of discussions. Following the logical framework introduced earlier, these data were analyzed in a step-wise fashion.

First, the essays were shuffled and given to an advanced philosophy graduate student for blind grading, and a dependent t-test was used to investigate pre-post differences. The statistical test was significant ($p < .001$, df. $= 28$). This finding provided evidence that the essays were indeed sensitive to <u>something</u>.

Second, to help ensure that this <u>something</u> was Knowledge and Reasoning, one student's pre- and post-tests (rated 1.0 and 3.0 respectively, on a four point scale with half-point gradations) were selected for further analysis. Regarding Knowledge, the student identified fewer and less pertinent issues in the pre-test than in the post-test. In addition, virtually none of the vocabulary of ethics was employed in the pre-test, whereas, in the post-test, notions such as

"parentalism", "autonomy", and "rights-based framework" were used to ferret out the issues and organize the discussion. Regarding Reasoning, the pre-test consisted of a string of assertions and a statement of a position. What little argument was present amounted to an uncritical appeal to conventional behavior. By contrast, in the post-test the student listed several alternatives (including the position taken in the pre-test), critically evaluated the alternatives, rejected them, and then defended an additional alternative against objections she anticipated. (See Howe, 1982, for the actual tests and a more thorough discussion of the differences between them.)

In conjunction with the statistical test, the analysis of the paired pre-post essays supported the existence of the constructs Knowledge and Reasoning, as well as the use of essays as viable means of measurement. It also demonstrated the virtues of multiple indicators, since the statistical test by itself was not very rich in meaning with respect to the explicit course goals (a point made by the instructors that helped motivate the more fine-grained analysis) and was subject to multiple interpretations.

Third, confirming (or disconfirming) evidence for Knowledge and Reasoning was sought in the observational data, and evidence supporting the constructs Knowledge and Reasoning (paralleling the evidence for these constructs in the pre-post essays) was detectable. The sensitivity of direct observation to Knowledge and Reasoning is

demonstrated below by a comparison of the first and last meetings of the Ethics in Nursing course in which evidence for Knowledge and Reasoning was not and was present respectively. The existence of detectable differences between the two class meetings implies perforce that Knowledge and Reasoning are measurable via direct observation and, because the observational findings may be added to the evidence from essays, that the basic functional model yields warranted conclusions.

When confronted with the pre-test (i.e., Margaret's dilemma) in the first session, students asked the following kinds of questions to "clarify" the assignment:

"Isn't that in the physician's medical category?"

"Isn't it allowable to say, 'You're seriously ill'?"

"I've never been in a hospital [qua nurse]. What is done?"

Compare these three questions to excerpts from a discussion which occurred in the eighth meeting concerning the case "Working Extra Hours" (Benjamin and Curtis, 1981, p. 122). The case involves the issue of the degree to which nurses and nursing students should volunteer their professional services. One nurse in particular has been volunteering a good deal of her time to the hospital in which she works, and this is resented by her co-workers. (Codes are indicated in brackets to facilitate subsequent discussion of this example: {1} = challenging proffered views, {2} = offering counter-examples, {3} = requesting clarifications, and {4} = defending one's position. Codings

follow the remark to which they apply.)

Student 1: Do you think that others would care if you took more hours at a college situation? It seems to me that it's kind of different than a work situation. In what ways are they alike? {3}

Nurse: Well, she is working these extra hours, and saying it's to learn these things so that she can be a better nurse. And I think she should be applauded for her efforts and encouraged in her efforts so that she can be... But I paid my dues. Why should I have to continue to do more? Would you give it away free to the hospital? I think that's the issue...

Student 2: I don't want to be assigned to a job that's going to require certain hours of me. I just want to volunteer my time to something else. Does that mean that I am being less than professional? It's free--You are getting my professional services free. {2}

Philosopher: What possible difference is there? Let's say you volunteer to the Hospice, say, here in town. Everyone in the Hospice group here in town is a volunteer. One of the things that leads to the friction is that she does this stuff as a volunteer and every one else is getting paid. That's the thing.

Nurse: And when she leaves, then unless somebody else comes along and volunteers, the work isn't going to be done without a lot of specially trained people being pulled into this, giving of themselves for free. In some sense it becomes part of their job.

Student 2: Could it be more justified then if she did go and ask for pay for this work? {3}

Nurse: If she were paid for the job, then the other students would feel it was--if the job had money attached to it. A woman who had other obligations and could only work 40 hours a week could just as easily work those hours doing that kind of job. Once it becomes paid and part of the job description for the nurses on that unit, it can be held by more than just the person who is able to volunteer. So then it would be a problem among the administration and staff.

Philosopher: Part of the problem here is what the codes say and all that. What are the limitations of your job or role as a nurse?

Student 3: This reminds me--do you remember last week? Someone asked what would you do at a cocktail party if someone you know comes up to you and wants advice on something. She [the nurse] said, "Well, if somebody approached me looking at me as a nurse, then I would refer them to somebody else. I wouldn't give out free information." You [referring to the nurse] were basing it on economic growth and you wanted nurses to be recognized as professionals and that your advice is valuable, so you weren't willing to give that out for nothing. And I thought about that for

awhile ... It all comes down to how everybody perceives their role as a nurse differently, and I thought about what was said--and you are perceiving it as an 8 to 5 job. And I do not accept that. So for me, you know, for myself, I don't see it as an 8 to 5 job, and so therefore I don't see anything wrong with it if you want to take extra courses or work extra hours ...{4} With your view [the nurse's] you are restricted there if you want to obtain more knowledge. {1}

Philosopher: The question is this: I want to know whether doing the things that you are talking about are things that you are permitted to do but don't have to, or, on your conception of nursing, these are things such that if you could do them and didn't, you would be neglecting you responsibility. Do you see what I mean? Nursing, after all, is doing things for people. Using your specialized knowledge and abilities without getting paid or anything like that--Is that a gift or is it a responsibility?

.

.

.

.

Student 3: I'm not in nursing for the money, so I can't base anything on economic reasons. I am not pushing for better pay... The thing too is that me being a nurse--that's part of you. You don't turn around and stop being a nurse. {4} I just got an awful lot of free advice from my doctor too, and, you know, it's very hard to see anything wrong with that. {2}

Nurse: But he's already your provider. That relationship was already established. He didn't go across a crowded room and...

Student 3: No, but on the other hand, while he is sitting there eating his hamburg he really didn't have to talk to me about my medical problems. {4}

Philosopher: He already has that relationship with you. He is your physician. I really think you have to think about, you know, just where you're drawing lines on what. If you drive by and see an accident, are you going to stop? When people come up and identify you as nurses, when do you give it away? I mean there's a difference between emergencies and non-emergencies. Likewise, there is a difference between strangers and those with whom you already have a relationship.

.

.

.

.

Student 4: Your position [the nurse's] to me seems to contradict what you said about self-determination and autonomy. {1} Maybe you can explain to me how this could be self-determination. {3} I think it's important to have self-determination...

The differences between the discussion in the initial meeting and the later one mirror the differences in the pre- post-tests. The vocabulary of the course is employed by the students, e.g., "autonomy" and "self-determination", in the later discussion, and there are also detectable changes in the way students go about the task of ethical inquiry. Rather than pumping the instructors for the **answer**, the students take an active role in a genuine Socratic exchange: they challenge proffered views, offer counter-examples, request clarifications, and defend (rather than assert) their positions (see codings accompanying the excerpt from the later meeting for specific illustrations). The use of appropriate vocabulary may be taken as a measure of Knowledge, and participation in the Socratic method may be taken as a measure of Reasoning (see Chapter II for the list of criteria associated with Reasoning). One may conclude that direct observation is sensitive to Knowledge and Reasoning.

This finding may be combined with the findings from the two kinds of analyses of essays. Taken together, the results of the three kinds of analyses strongly support that the constructs Knowledge and Reasoning can be measured and, per force, that they exist. The results of the analyses also support the value of the basic functional model. In short, the three kinds of analyses combine to support the arguments of Chapter II that Knowledge and Reasoning are appropriate goals and evaluative criteria for medical ethics

teaching, and the arguments of Chapter III in defense of a functional evaluation approach.

Appreciation. With good evidence for cognitive effects in hand, a start was already underway on the question of Appreciation. That is, because Appreciation has both cognitive and attitudinal dimensions and the former had already been established, positive attitudes about the course would support the existence of the construct Appreciation and that it is measurable.

On the other hand, Appreciation is a more useful construct if it can be inferred in the absence of independent evidence for cognitive effects. One cannot be assured ahead of time that cognitive effects will be demonstrated. Investigations of cognitive effects typically demand significantly more resources than measurements of attitudes and may be otherwise unworkable in many contexts. If evidence for Appreciation is relatively self-standing, then the case for Appreciation (versus wants or likes) is that much stronger when such self-standing evidence is combined with independent evidence for cognitive effects.

In the Ethics in Nursing study, the measurement of Appreciation was distinguished as much as possible from the measurement of cognitive effects. As described in Chapter II, Appreciation of medical ethics involves a positive attitude about the right objects (e.g., the value of cognitive ethical inquiry). Consistent with this focus, the mailed student evaluation form was constructed to emphasize

issues of particular relevance to medical ethics teaching. The analysis of free responses on the SIR forms also emphasized issues of particular relevance to medical ethics. Both kinds of student evaluation data were augmented by data from direct observations.

Selected results from the mailed evaluation form are displayed in Table 1. (The response options for the items were: strongly agree, agree, neither agree nor disagree, disagree, and strongly disagree, and the values 5, 4, 3, 2, and 1 respectively were assigned to these responses.)

Table 1. Selected Results From Ethics in Nursing Course Mailed Evaluation Form (Response rate = 23 of 32 or 72%)

Statement	Mean	Range
1. The course improved my ability to see the complexity of moral problems which nurses face.	4.7	4-5
2. The course helped me develop a framework or basis for my moral positions.	4.0	3-5
3. The course helped me see the importance of giving reasons and careful arguments for my moral beliefs and decisions.	4.5	4-5
4. The topics discussed and the written assignments were relevant to nursing.	4.7	4-5
5. The cases discussed were realistic enough to bring out the emotional side of moral problems.	4.6	3-5
6. I feel more confident about recognizing and dealing with moral problems because I have had this course.	4.3	3-5
7. All medical professionals should study moral problems in medicine in a similar way.	4.7	3-5
8. Multiple-choice or true-false exams would have been a better way of grading than essays.	1.4	1-3

Items 1-3 are most relevant to the cognitive dimension of Appreciation; items 4-7 are most relevant to Appreciation's attitudinal dimension. The relatively high means on items 1-7 support Appreciation as a measurable effect of the course. The responses support Appreciation rather than mere wants or likes because the items are keyed to the "right objects" (i.e., the value of cognitive ethical inquiry and its relevance). Although the halo effect may well be at work regarding items 1-7, the mean and range for item 8 eliminates the hypothesis that students settled on the upper end of the response scale and, having made this decision, unthinkingly responded to the remainder of the items.

The free responses on the SIR forms provide collateral evidence for Appreciation as a measurable construct (response rate = 32 of 32 or 100%). In this case, students zeroed in on the "right object" with no prompting from pre-formulated items. For example, 15 of 32 respondents (47%) specifically mentioned the relevance of the course. Their comments included:

> What I liked most about this course was that it was in direct relation to my field...

> This class was most interesting as we as nurses are going to encounter these dilemmas.

> I have really gained a heightened sense of awareness of ethical considerations of situations I face daily at work.

Paralleling these general comments, 13 (41%) praised the topics. Two examples of these are:

> One thing I liked most was that many topics were brought up that are not brought up in other classes. This has a more realistic approach to it--these are things that can and do happen...

> I felt the topics reached into many different facets of the profession. This exposure has helped me to formulate ideas that I never knew about.

Finally, 6 (19%) volunteered that the course was important enough to be required of all nursing students. One student wrote, "This class was just great and I wish so much that it was required. It's _insane_ that it's not."

The third kind of data supporting Appreciation as a measurable effect was direct observation of discussions. Like measurement of cognitive effects, direct observation of Appreciation is an independent criterion in the sense that it does not depend on the perceptions of students. Direct observation of the discussions indicated that student interest in the discussions was generally high. In the half-session in which "Working Extra Hours" was discussed, for instance, 28 of the 32 students enrolled were present and, of these, 22 participated directly in discussion. Other indications of Appreciation were observed attentiveness and interest during discussion and the tendency of the students to talk among themselves and with the instructors before and after the formal meeting time and during the breaks between the half-sessions.

The three kinds of data--the mailed evaluation forms, SIR forms, and direct observation--provide independent evidence for Appreciation as a measurable outcome of the Ethics in Nursing course. Independent evidence for

cognitive effects from the essays and observations further supports the existence of the construct Appreciation and its measurability.

Evaluation Impact

The previous section on results provided affirmative answers (amplified below in the concluding section) to two of the meta evaluation issues used to frame the discussion of the Ethics in Nursing example: (1) Knowledge, Reasoning, and Appreciation can be measured and (2) can work together within a functional evaluation approach to yield warranted conclusions. This section briefly addresses the third meta evaluation question, namely, whether empirical results obtained from the application of the functional evaluation approach can affect practice.

Evaluation research in general is plagued by laments about poor utilization of results. Chapter III argued that two of the causes of this problem are the use of inappropriate vocabulary (i.e., the vocabulary of moral psychology) and an associated over-emphasis on methodological rigor (i.e., "scientistic" evaluation) that cause the "questions of interest" for medical ethics teaching to be lost. Hence, it is germane that the evaluation methods employed in the nursing ethics course generated results that were interesting and useful to relevant audiences.

The most immediate impact of the evaluation was that it convinced the instructors of the nursing ethics course that

an effective course had been developed; the course was taught in subsequent years in virtually the same way. The evaluation also had a more diffuse impact. Its results were communicated to broader audiences of individuals interested in ethics teaching. One version was directed at an audience of philosophy teachers (Howe, 1982), and was designed to show that ethics teaching is evaluable in terms of methods and criteria acceptable and familiar to ethics instructors. Other versions were presented in conference and workshop formats to a more limited audience of those interested in nursing ethics in particular. These versions were designed to promote the design and content of the course and to suggest means of initiating and evaluating similar courses.

Conclusion

The Ethics in Nursing example entertained three primary meta evaluation questions: (1) Can the goals of Knowledge, Reasoning, and Appreciation be measured? (2) Can the basic functional model for medical ethics courses yield warranted conclusions? and (3) Can the empirical results obtained by employing the model affect the practice of medical ethics teaching? The answer to all three questions is "yes".

(1) Knowledge and Reasoning were measured using essays and observation. The statistically significant difference between the pre- and post-test scores provided one piece of evidence that Knowledge and Reasoning are valid and are measurable. Against the scientistic fact-value distinction, the grader distinguished pre- from post-tests blindly, in

terms of cognitive criteria such as the use of appropriate
vocabulary and quality of argumentation. This speaks for
objective cognitive standards by itself and was further
supported by multiple indicators: the analysis of one pair
of essays in terms of Knowledge and Reasoning and the
analysis of observational data supported the same
conclusion. One may conclude that (contrary to scientism)
recognizable (objective) standards for detecting ethical
Knowledge and for distinguishing better from poorer ethical
Reasoning exist and are measurable.

Appreciation was measured using direct observation and
student evaluation forms. The observational data provided
evidence for the construct Appreciation in terms of criteria
like interest, participation, and regular attendance. The
student evaluation forms provided more focused evidence for
Appreciation, and further support for Appreciation (versus
mere wanting or liking) followed from the independent essay
testing evidence that indicated cognitive effects
concomitant with the positive attitudes exhibited in the
student evaluation forms.

(2) The very same evidence and arguments that support
the measurability of Knowledge, Reasoning, and Appreciation
support the warrant of the conclusions drawn from the Ethics
in Nursing Example. That is, evidence that the various means
of data collection--essays, observations, and student
evaluation forms--are sensitive to Knowledge, Reasoning, and
Appreciation is also evidence that those three things exist,

to be measured.

(3) Against the scientistic demand for psychologically grounded criteria and measures, the customary reason for this demand--that it is the only way to generate objective, L-fallible results--had no force. In addition to the problems discussed in Chapter II (that scientistic criteria hypostatize Reasoning) and Chapter III (that scientistic criteria and methods are invalid, impracticable, or both) a method that employed non-scientistic criteria (Knowledge, Reasoning, and Appreciation) and measures (essay tests, questionnaires, and observations) generated results sufficiently L-fallible to be credible to the relevant audiences.

Example 2: Focal Problems (1982)

Course Description

Focal Problems is a three-course sequence required of all second-year medical students enrolled in the Track I[3] curricular option at Michigan State's College of Human Medicine. In fall 1981, support from the National Fund for Medical Education and the National Endowment for the Humanities made it possible to revise the courses to incorporate decision analysis and ethics.

Each of the courses meets for a ten-week term. The format combines one hour of lecture per week and two hours of small group discussion. The discussion groups consist of from 8 to 12 students and are each staffed by two

preceptors, one physician and one faculty member from the
behavioral sciences or the humanities. Discussion in the
groups focuses on applying the material presented in lecture
to paper-cases designed to approximate medical
problem-solving situations. The first two terms emphasize
decision analysis (see Howe et al., 1984, for a fuller
description); the third emphasizes ethics and is the concern
here.

Four cases, chosen to coincide with students' course
work in basic science and to create a logical progression of
issues, comprised the ethics course: (1) a four-year old
rendered brain dead by smoke inhalation, (2) Karen Quinlan,
who though not legally dead, exists in a "persistent
vegetative state", (3) an apparently competent burn victim
who wishes to be allowed to die, and (4) an ill, possibly
demented elderly woman who is a "social admission" to a
hospital. The bio-ethical issues raised in the cases
included the following: the distinctions between brain
death, poor prognosis, and persistent vegetative state; how
courts impinge on bio-ethical issues; refusal of medical
treatment by competent versus incompetent patients; the
conflict between physician paternalism and patient autonomy;
the conditions under which patients may be declared
incompetent and the relationship between medical criteria
and moral considerations involved in establishing such
conditions; and access to medical care for the indigent in
light of costs and aims of medical care.

Lectures keyed to the cases were given by philosophers
from Michigan State University's Medical Humanities Program
(MHP). Except for one MHP philosopher who volunteered,
preceptors for the small groups were assigned solely on the
basis of availability. Students were evaluated on the basis
of two midterm examinations and a final. Each exam was
composed exclusively of fixed-response items, the customary
examination format in the college.

Faculty development consisted of two two-hour seminars
composed of a general introduction to the nature and aims of
medical ethics teaching and practice in discussing the
cases. Preceptors were also provided with "leader's guides"
to assist them in conducting the discussion groups.

Evaluation Purposes

The evaluation of this course, like the evaluation of
the ethics in nursing course, aimed to generate findings
useful for informing improvements. Various features of the
Focal Problems course, however, limited the degree to which
the methods and findings of the ethics in nursing evaluation
were applicable. It was required versus elective, it was
part of medical school versus undergraduate curriculum, and
it was staffed in large part by inexperienced conscripts
versus enthusiastic devotees. Added to this, the course was
was part of a larger project, "Ethics in the Core
Curriculum",[4] involving Michigan State University's two
colleges of medicine and its college of nursing. The
project's aim was to integrate ethical issues

into existing educational activities wherever feasible, employing philosophers as curriculum developers but regular faculty as teachers. The Focal Problems course approximated this general approach and thus provided a relatively controlled and circumscribed context to test the integration model, from which results could be generated quickly and applied as appropriate to the more diffuse and inchoate activities of the project.

Evaluation was also intended to assess the view that integration using regular clinical faculty is the way to demonstrate to health professions students the relevance and value of serious study of medical ethics. Given the pivotal position of the clinical faculty--linchpins between the material and the students--the evaluation of the Focal Problems course was two-tiered. One substantive purpose of the evaluation was assessing the teaching performance of the non-philosopher faculty, especially physicians, since their acumen and support (Knowledge, Reasoning, and Appreciation) was relevant to the warrant of the integration model and constituted in large measure the execution of the course.

In the process of investigating substantive questions about the Focal Problems course and the integration model, three meta evaluation questions were also pursued: (1) Can fixed-response tests supplant essay tests as a measure of Knowledge and Reasoning? (2) Can the basic functional model defended in this chapter detect evidence for indirect goals? and (3) Can the model be adapted to the medical school

context and yield useful results? Consistent with the
emphasis in this chapter on meta evaluation issues, the
remainder of the discussion of the Focal Problems example
will focus on these three meta evaluation questions. The
subsequent "evaluation design" and "evaluation results"
sections focuses on framing and answering the first two
questions, the "evaluation impact" section answers the
third. The concluding section reviews the major meta
evaluation lessons this example provides.

Evaluation Design

Data were collected with two evaluation forms,
semi-structured interviews with the preceptors, direct
observation of lectures and discussion groups, and a
fixed-response cognitive performance test. These data
collection techniques were keyed to Knowledge, Reasoning,
and Attitudes in ways that should be familiar to the reader
by this time. Observation and interviews were applied to
the "indirect goals", in addition to their uses in the
previous example. Also, interviews of preceptors were

	Knowledge	Reasoning	Attitudes	Indirect Goals
fixed-response tests	x	x		
observation	x	x	x	x
evaluation forms				
mid-term			x	
end-of-term			x	
interviews	x	x	x	x

Figure 5. The Relationships Among Data Collection Techniques and
Constructs for Focal Problems (1982)

to assess students' Knowledge and Reasoning. Figure 5 summarizes the various purposes of the data collection methods used in the Focal Problems evaluation design.

Observations had both students and preceptors as targets. As intimated above, the preceptors themselves were one group at which the curriculum was aimed, i.e., they had to be brought up to speed in Knowledge, Reasoning, and Appreciation before they could be expected to impart these to students. Resources did not permit earlier and later observations of the small group discussions; an investigation of progress within a discussion group (on the model of the nursing course) was thus not possible. Each small group was observed only once.

An O X O design was again employed for the pre-post cognitive testing. Unlike the nursing course, however, a fixed-response examination was used rather than short essays. Several considerations led to this decision. First, medical students are tested exclusively by this means in the remainder of the curriculum, and the course designers desired that the ethics course be as similar as possible in format to other course work. In addition to serving as a measure of entering skill and knowledge, the cognitive test also introduced students to the nature of the testing that would be employed in the course. Although fixed-response examination formats were the norm, the examinations designed for the Focal Problems ethics course required a good deal more by way of inferences (versus recognition or recall)

than students were accustomed to because they were designed to be sensitive to Reasoning as well as Knowledge. Second, given the multiple small group format (6 groups in all), the logistics of grading essays posed problems. Third, it was doubted that preceptors had the skills necessary to grade essays reliably. Unreliable grading creates serious problems by itself, and in the context of medical school, where ethics is often viewed as "soft" and "a matter of opinion", unreliable grading is taken as evidence that these claims are true.

Interviews of instructors loomed large by contrast to the ethics in nursing course. The interviews were designed to assess the performance and receptiveness of preceptors, gather formative information based on their perspective, and obtain their estimation of students' Knowledge, Reasoning, and Appreciation.

Evaluation Results

The discussion of results is divided into sections on Appreciation, Knowledge and Reasoning, and indirect goals.

The section on Appreciation is not directly applied to the three meta evaluation questions that frame this example. Instead, the discussion of Appreciation reinforces the argument of the nursing ethics example that Appreciation is measurable and illustrates how Appreciation may be extended to medical school faculty. These two issues are discussed with an eye toward the subsequent "impact" section of this example (Focal Problems, 1982) and an eye toward the

evaluation design in the next example (Focal Problems, 1983).

The section on Knowledge and Reasoning emphasizes the meta evaluation question of whether fixed-response tests can supplant essay tests as a measure of Knowledge and Reasoning. The section also illustrates the value of combining other indicators of Knowledge and Reasoning, such as observations and interviews, with tests. Knowledge and Reasoning, like Appreciation, are also extended to medical school faculty.

The section on indirect goals is keyed to the meta evaluation question of whether the variant of the basic functional model employed in Focal Problems can be sensitive to indirect goals.

Appreciation. The course was positively received by the preceptors. Of the 13 who participated, 11 were interviewed (one was unavailable and the other was an MHP philosopher). Each of the respondents believed the course was highly valuable and that such experiences are important for an adequate medical education. Although two had some reservations about the course being the best approach, all expressed an interest in teaching it again. These findings are germane to Appreciation. Below are two excerpted comments that illustrate how preceptors' praise focuses on the "right object" (i.e., the value of cognitive ethical inquiry).

> Psychiatrist: It's kind of reassuring to know these things are being discussed versus merely assimilated as they were when I was
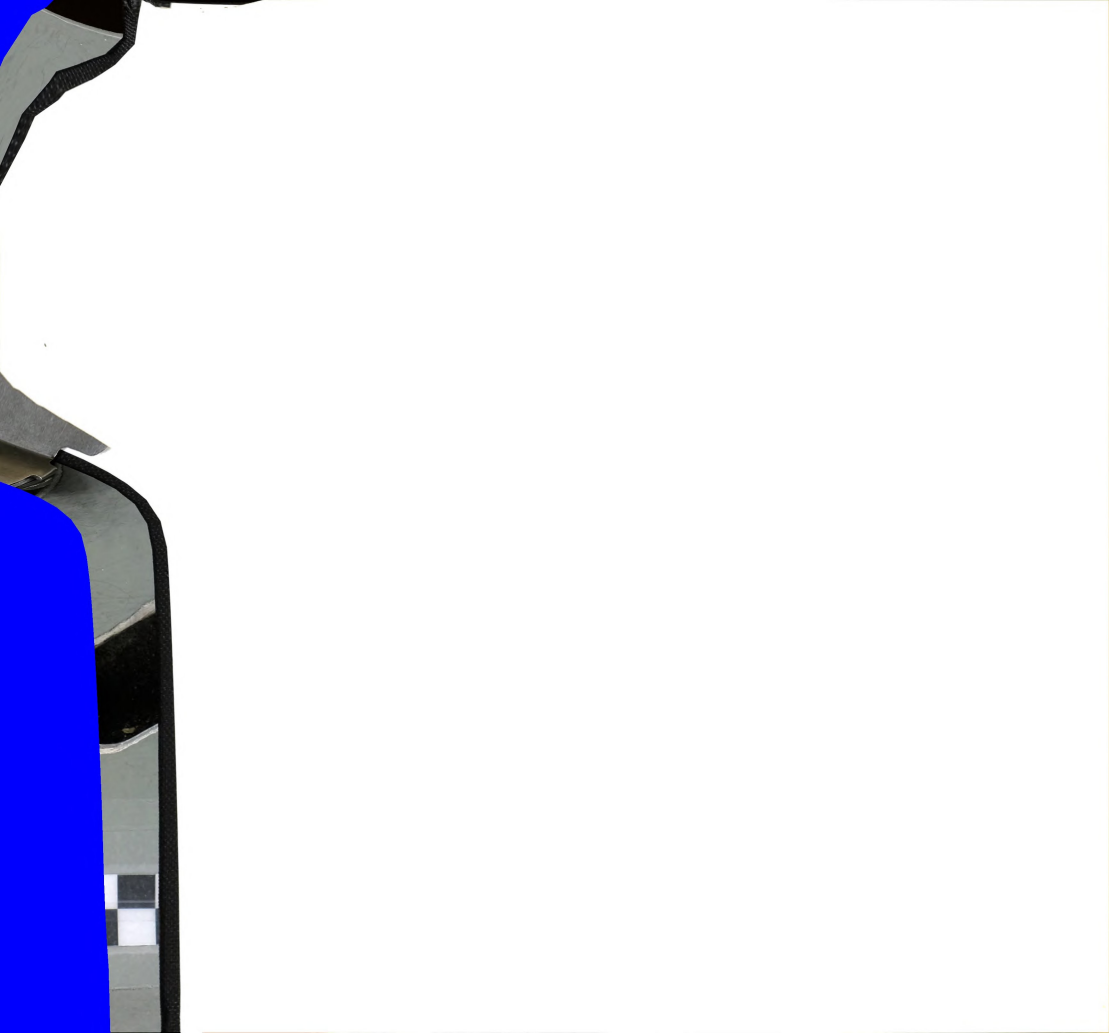
in medical school. Not that we practiced unethically, but we freelanced and didn't think it through and discuss it as much as we should have.

Internist: Medicine is not certain. There's an open gate. I've got this 25 year old girl with an uncertain diagnosis but a suspected malignancy. What do you tell her? Ethics is no different from any other area of medicine. It isn't always clear what to do. You have to use clinical judgment, the circumstances, literature, precedent, etc.

The materials for the course were rated good, very good, or excellent by 10 of the eleven preceptors; 1 rated them average. The "leaders' guides" were judged extremely helpful and complete, but a number of preceptors urged that the practice of distributing them to the students be changed (this urging is relevant to the later discussion of preceptors' Knowledge and Reasoning).

Results of an evaluation form distributed at the end of the term showed that the students also received the course well. Seventy percent of the students said that it is "quite important" for physicians to be skilled in dealing with ethical problems, and all the rest called it "important". Over three quarters of the respondents believed the course would help them deal with such problems in the future. These positive attitudes apply to the "right objects"--namely, the relevance and value of careful study of medical ethics--and thus support Appreciation as opposed to mere wanting or liking.

On the negative side, some made the familiar claim that ethics is "unteachable". A significant minority (6) criticized what they believed were biased materials, lectures, and tests. A few (3) believed themselves no match

for the lecturers and called for readings and positions
representing conservative and religiously grounded
viewpoints. Testing received the greatest criticism.
Nearly half of the students thought the exams were too
difficult, and 17% believed they were otherwise
inappropriate or irrelevant.

An evaluation form distributed near the middle of the
term foreshadowed these results: the students generally
viewed the course as worthwhile but grumbled about the
tests. Moreover, preceptors testified to this state of
affairs in the interviews and it was also evident from
observations. Despite the acrimony over testing, however,
students were genuinely interested in and engaged by the
course, both in discussion and lecture.

In summary, at the level of meta evaluation, the
construct Appreciation could be applied to both students and
preceptors. At the substantive level of evaluation
findings, preceptors exhibited a somewhat higher level of
Appreciation (or at least approval) than students.

Knowledge and Reasoning. The Focal Problems course
showed that interview data could be applied to the
constructs of Knowledge and Reasoning. The reasons that
were given by preceptors for withholding the "leaders'
guides" from students are relevant to preceptors' levels of
Knowledge and Reasoning; they gave two reasons: (1)
Providing students with the guides encourages them to "look
up the answers" and thus inhibits careful thought and

reflection. (2) It disarms preceptors--one remarked, "Brody and Miller [the curriculum designers] have exhausted the issues at our level of sophistication in the leader's guides". These concerns suggest that the preceptors lacked the necessary skills (loosely, Knowledge and Reasoning) to perform optimally. The philosopher-preceptor had no similar misgivings about distributing the "leaders' guides" to students.

Other remarks made within the interviews added further evidence about preceptors' ability to perform within the discussion groups. Most preceptors reported that they sometimes groped in discussion group and ran out of things to say--problems they believed philosophers would not encounter. One reported he sometimes didn't know the right "philosophers' moves"; another, that the discussions were sometimes a bit too "touchy feely"; and, a third, that "the discussions were sometimes hard to keep on track". Direct observation confirmed that these difficulties did arise in all of the discussion groups except the one staffed with an MHP philosopher. This group was also rated higher than the others on student evaluation forms, as were the lectures given by philosophers.

Students' Knowledge and Reasoning was examined in three ways: a pre-post fixed-response cognitive test, observations of discussions, and preceptors' opinions elicited in interviews. Evidence regarding students' Knowledge and Reasoning was unclear. On the one hand, testimony from the

preceptors and observations indicated that students did gain in terms of use of the appropriate vocabulary, use of distinctions, consideration of alternatives and objections, and so forth, much as the nursing students had done in the previous example. On the other hand, and unlike the nursing example, progress was not evidenced on the basis of the pre-post testing. On the contrary, a dependent t-test was significant in the wrong direction (p < .01, t = -2.49, df = 56). Rather than mutually supporting Knowledge and Reasoning, as the multiple indicators (observation and essay testing) did in the nursing evaluation, the multiple indicators (observation, interviews, and fixed-response testing) conflicted in the Focal Problems evaluation.

It is highly unlikely that (as the statistical test suggests) students actually became worse in terms of Knowledge and Reasoning: such a conclusion is implausible on its face and is contradicted by the evidence from observations and interviews. These contradictory findings about Knowledge and Reasoning illustrate the value of multiple indicators. Without the data from observation and interviews, the testing results would be less easily dismissed. As one consultant to the "Ethics in the Core Curriculum" project (namely, Michael Scriven) remarked, "The multiple indicators saved [the evaluation]".

The mixed results about students' Knowledge and Reasoning led to a negative conclusion regarding the meta evaluation question of whether Knowledge and Reasoning
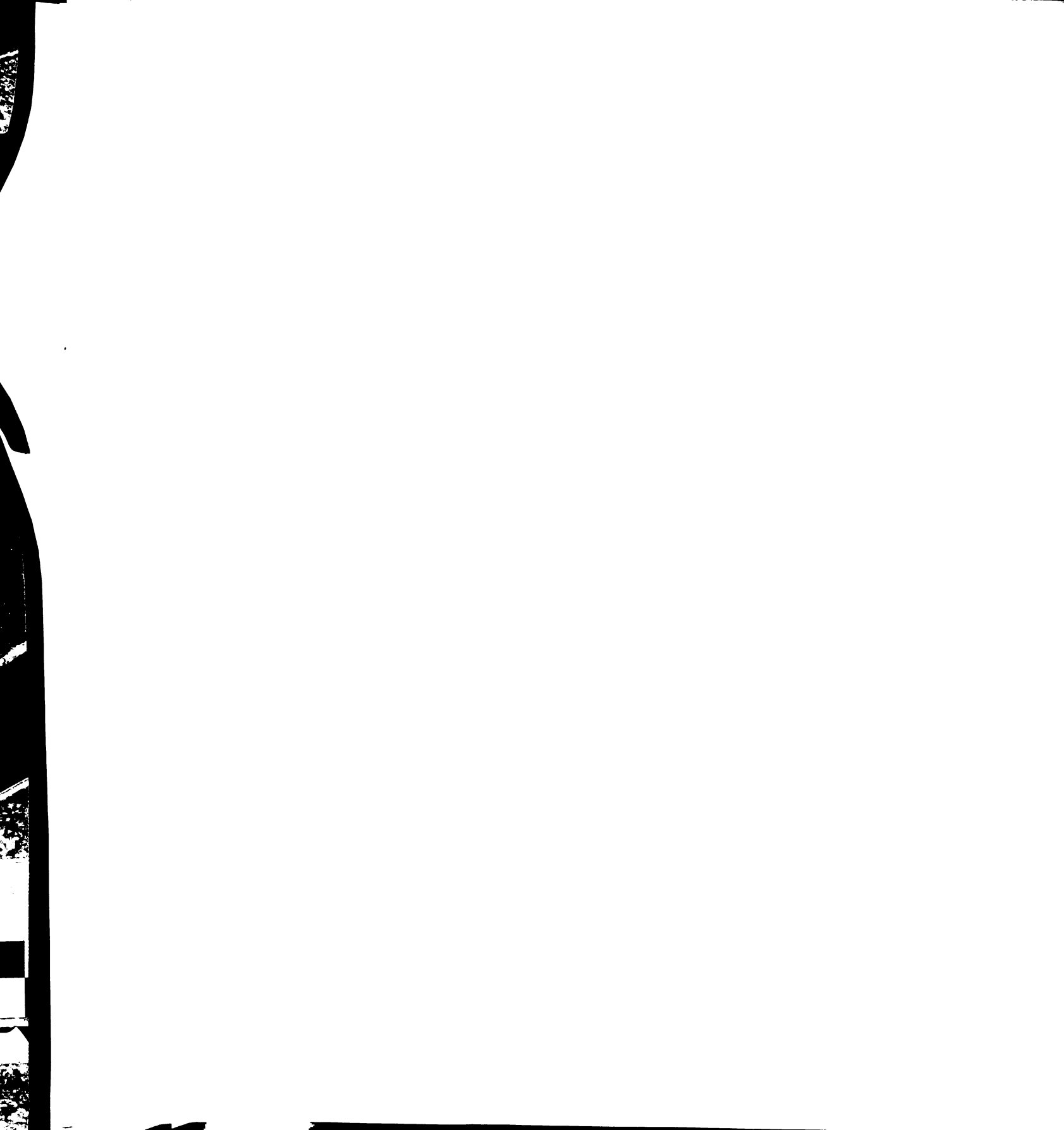
should be measured with fixed-response tests. The attempt
to measure Knowledge and Reasoning with a fixed-response
test in the Focal Problems evaluation failed, and, although
the test had virtues (described below), these were
outweighed by several disadvantages that probably cannot be
overcome in most educational contexts.

The exam consisted of 33 items and required less than a
half hour to complete, which rendered its .785 reliability
(KR-20) quite respectable. It was designed to be more
sensitive to particular content and experience than measures
such as James Rest's DIT, for instance, and data suggested
it possessed the desired characteristic (i.e., construct
validity). When given to groups of philosophy
undergraduates, medical students, preceptors, and philosophy
professors, it ranked them in the predicted order:
philosophy professors scored highest, next came preceptors,
then medical students, and finally undergraduates (means
equalled 29.75, 24.85, 24.08, and 21.6 respectively).
Finally, the negligible difference between medical students
and preceptors (24.08 versus 24.85) is consistent with the
difficulty preceptors experienced in leading the students in
discussion. (See Jones and Howe, 1983, for a more thorough
discussion of the instrument.)

Despite these virtues, the test proved inappropriate for
the circumstances. Indeed, it suffered from some of the
same shortcomings considered in connection with Kohlbergian
measures in Chapter III. It lacked face validity: students
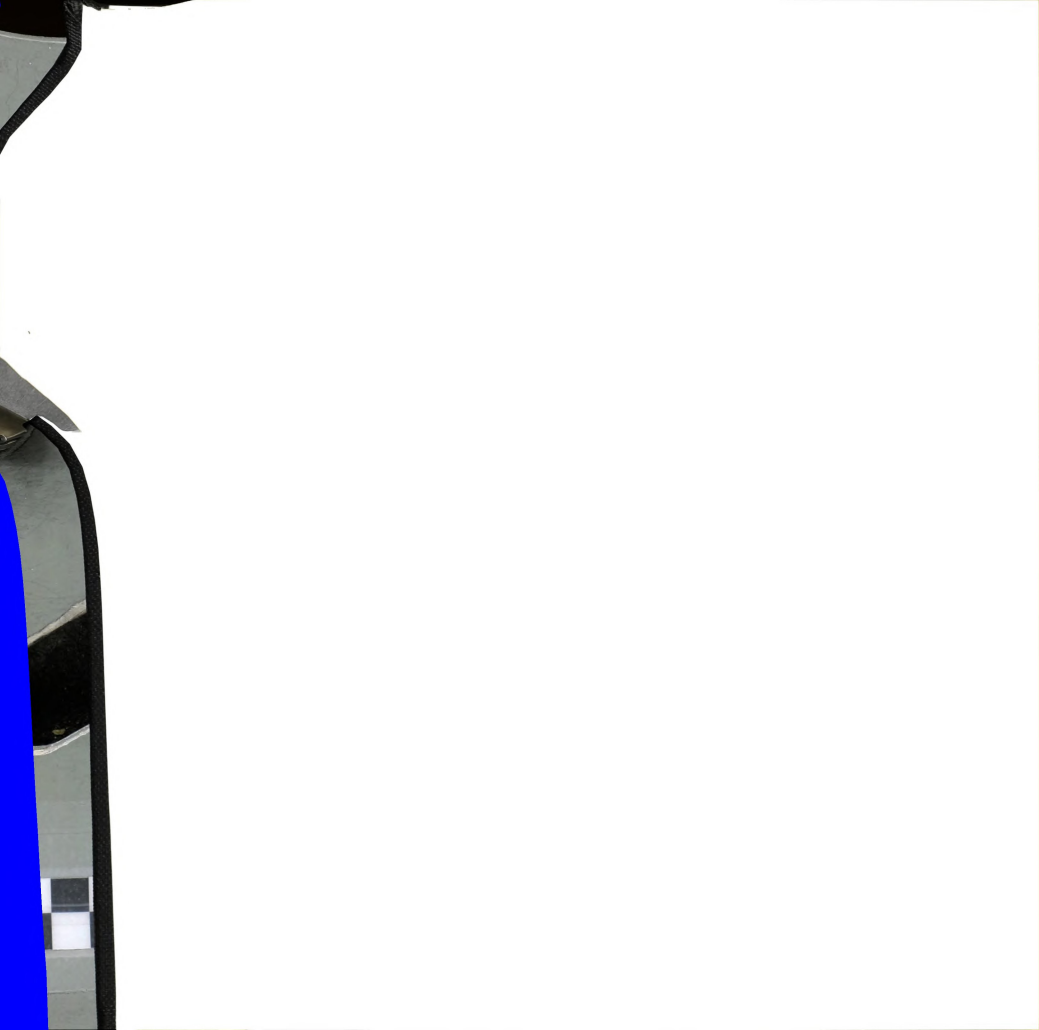
described it variously as "brutal", an "IQ test", a "logic
test", and a "reasoning test". The exams used throughout
the term and approximating the pre-post instrument were
similarly described and were, accordingly, ineffective for
feedback. In addition, the descriptions suggest that the
exams were too formal and hence unlikely to detect effects
which could be anticipated to result from a ten-week course.
This judgment was reinforced by the observations of some of
the preceptors as well as an outside consultant.

The testing experience in Focal Problems provides some
important practical lessons that count for functional and
against scientistic evaluation. It is conceivable (indeed,
quite likely) that the exclusive fixed-response testing
strategy could have been preserved in the name of rigorous
evaluation. Three important practical considerations,
however, argue against doing this. First, fixed-response
exams require considerable time and technical skill to
develop, especially when they are designed to measure
reasoning skills. Because they require so much time and
effort, it is wise to keep such exams "secure" (i.e., to not
return them to students), a practice that compromises
effective feedback to students on their performance.
Second, the results that are generated from such fixed
response exams are of questionable value. Because such
tests lack face validity, just what to make of results, even
if positive, is unclear to those who are interested but who
question the relevance of such tests. Third, such exams can
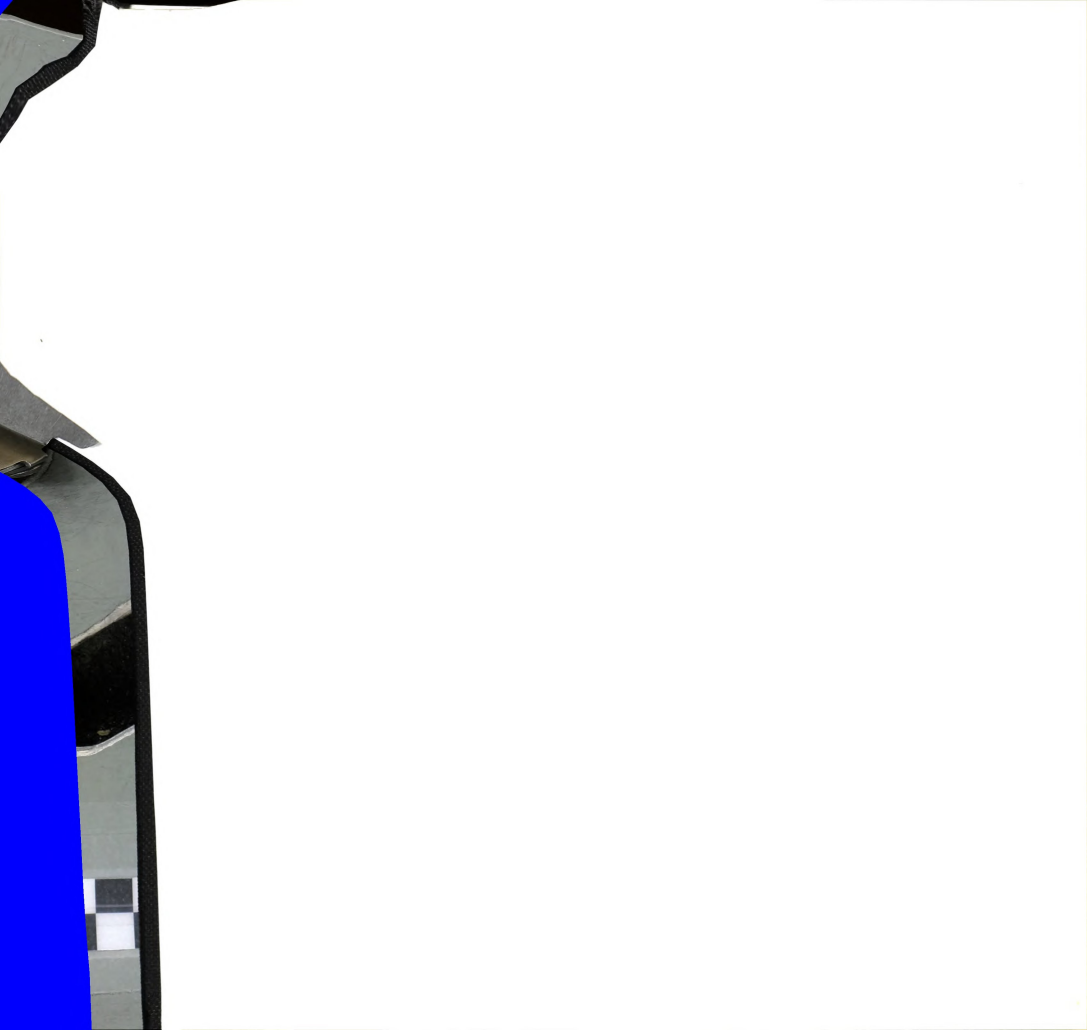
be obstructive. Using them in the name of rigorous evaluation findings can compromise the effectiveness of the program itself (as it did in the Focal Problems course by prompting considerable criticism and ill-will).

Indirect Goals. The second meta evaluation question entertained in this example is whether the design employed can be sensitive to the indirect goals of medical ethics teaching. The observations of the small groups and interviews of preceptors suggested that behaviors associated with indirect goals—goals of the "process" of medical ethics teaching—were evident within the Focal Problems discussion groups. Although the evaluation was in fact conducted in terms of criteria such as emotional involvement, interest, seriousness of purpose, and participation, the results may be recast in terms of Moral Regard, Empathy, Interpersonal Skills, and Courage.

In general, observations and reports from the preceptors within interviews indicated that students were stimulated to care (show Moral Regard) and to identify (Empathize) with the patients discussed in the cases. Students also showed the ability to interact with one another about the issues raised by the cases (Interpersonal Skills) and were willing to express their personal views, even if unpopular (Courage).

Two specific examples support these general impressions about indirect goals. The first involves "enhancing Interpersonal Skills" and "promoting Courage", and is based
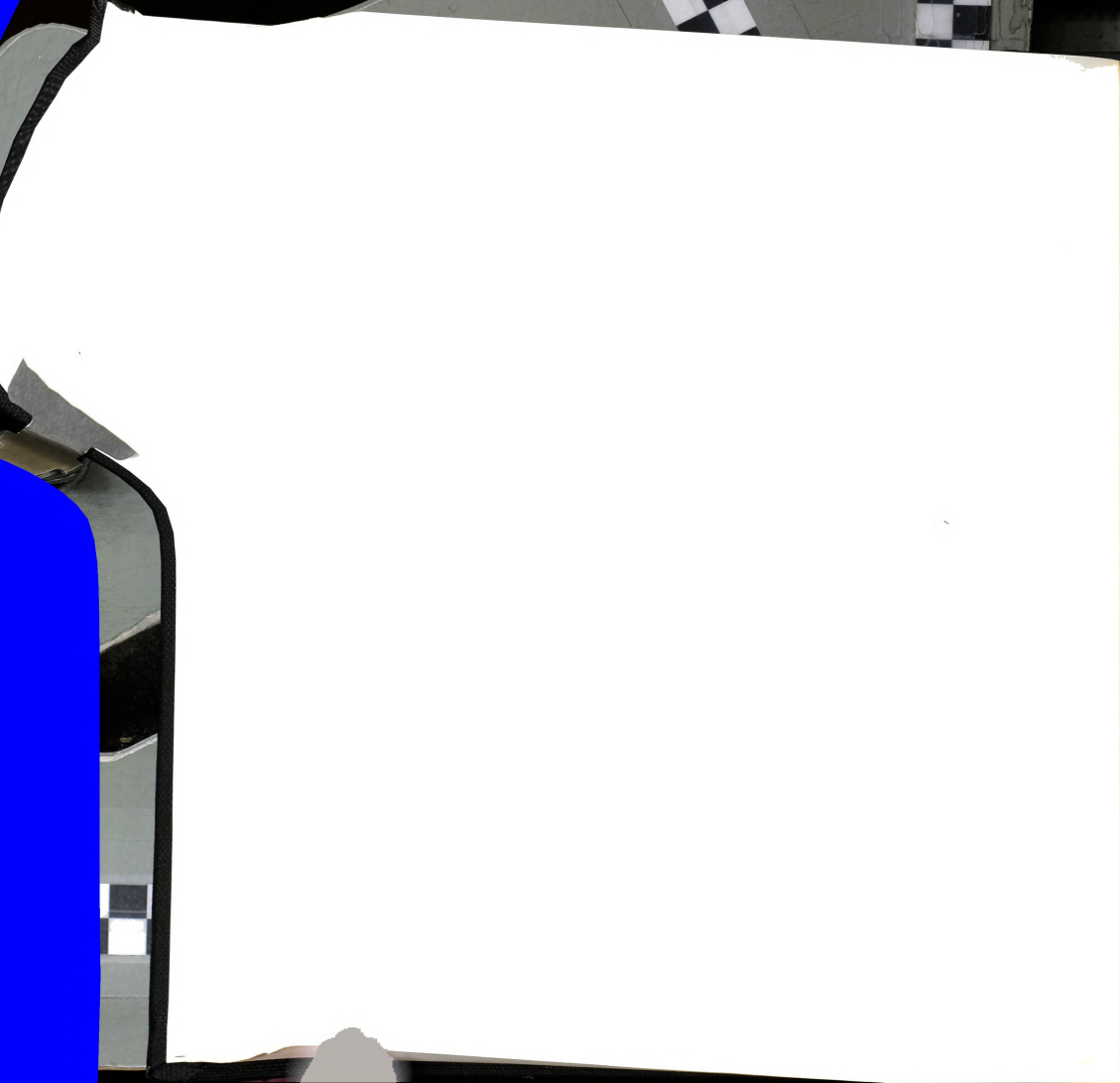
on interview data. One preceptor observed that the

importance of the issues to students as future physicians

plus the difficulty of resolving them (students couldn't

"look up the answers") required students to confront one

another about disagreements--something they were not

accustommed to doing. Initially, they became confused,

frustrated, and, at times, quarrelsome. The preceptor

remarked,

> Maybe this is not fun for the students. They're going to have to
> face this. It's frightening for them...if they don't get it, they
> become so confused and upset they don't even know what to ask for
> after awhile...Sometimes the discussions generated helplessness,
> depression, and anger and stimulated some unfair ways of arguing.
> Sometimes they were blocked by students' emotional makeups--for
> example, students with a fundamentalist background. Sometimes
> civility just went down the tubes.

The preceptor in question went on to claim that students

became better at "civility" as the term progressed.

Now, these data may be related to the goals of

"enhancing Interpersonal Skills" and "promoting Courage" in

the following way. Courage is required to express an

unpopular view. In so far as the discussion groups created

the context in which unpopular views are brought to the

surface, and in so far as the context encourages the

expression of such views, Courage is "promoted". At the

same time, because agreement (or compromise) must be

hammered out in this context of competing views,

Interpersonal Skills are "enhanced".

The second example involves the goals of "stimulating

Moral Regard" and "eliciting Empathy", and results from

observation and an informal interview with a preceptor. The
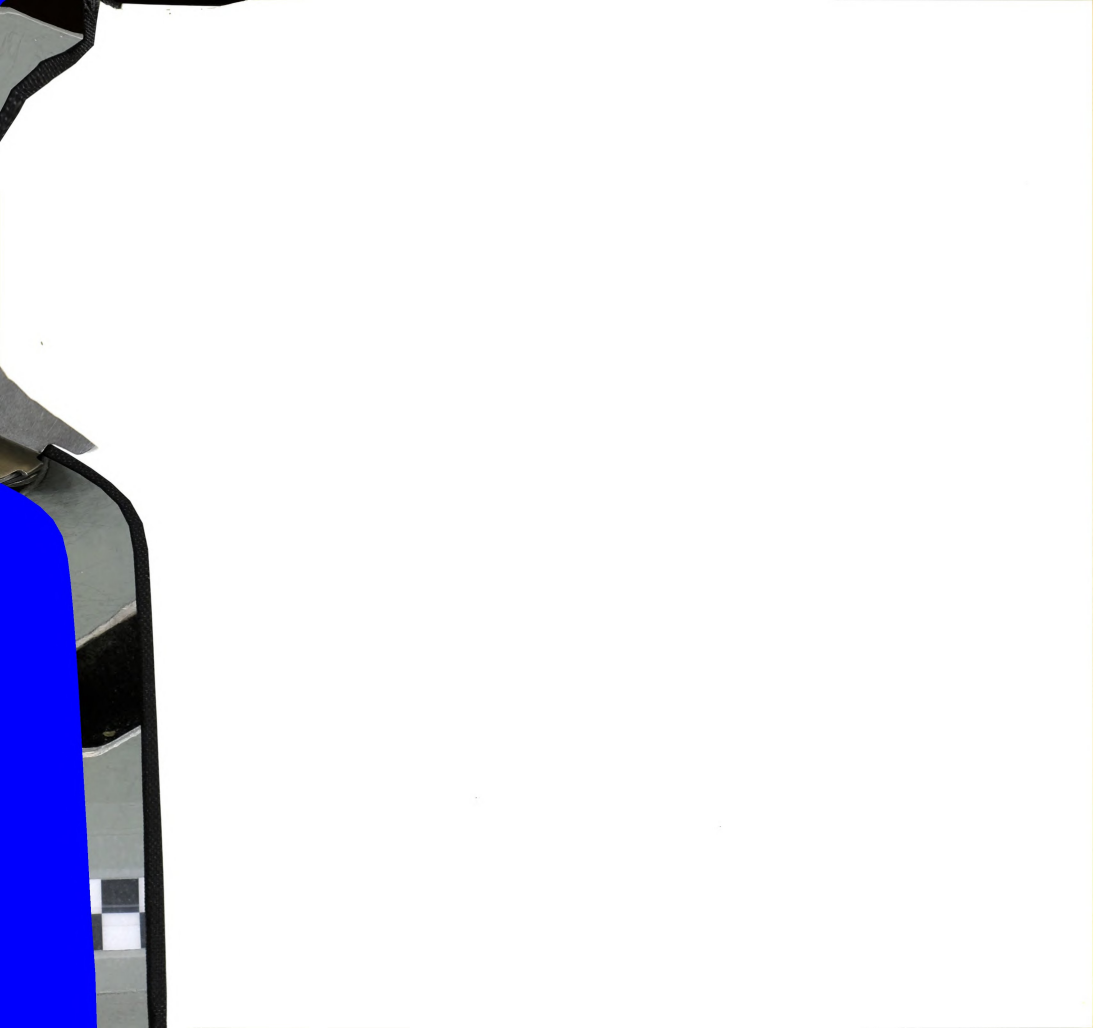
setting was the viewing of a videotape that provides a
particularly graphic depiction of the disfigurement and
painfulness of therapy for severe burn patients. The
primary ethical question at issue was whether the patient's
request (expressed in the videotape) to forego treatment and
to be allowed to die should be honored.

Following the showing of the videotape, students spent
about half an hour discussing therapy for burns. One
student, who had worked for several years as an orderly in a
burn center, reported to the group about how therapeutic
techniques had progressed since the videotape was made
(1975), how inept the attendants were, how outdated the
facilities were, and so on. Finally, one student asked
whether the prognosis for severely burned patients was
really much better at present and whether therapy was any
less painful. This question, to which a satisfactory answer
was not proffered, turned the discussion toward the
difficult cognitive issues raised by the case, for example,
the conditions under which patients may be judged
incompetent to make treatment decisions and the conditions
under which medical paternalism is justified.

The portion of the discussion most relevant to Moral
Regard and Empathy was the first half-hour. The outside
observer interpreted the discussion of the present state of
burn therapy as a lack interest by the students in the
ethical issues, and a preference for the technical, purely
medical side of the problem. By contrast, one of the
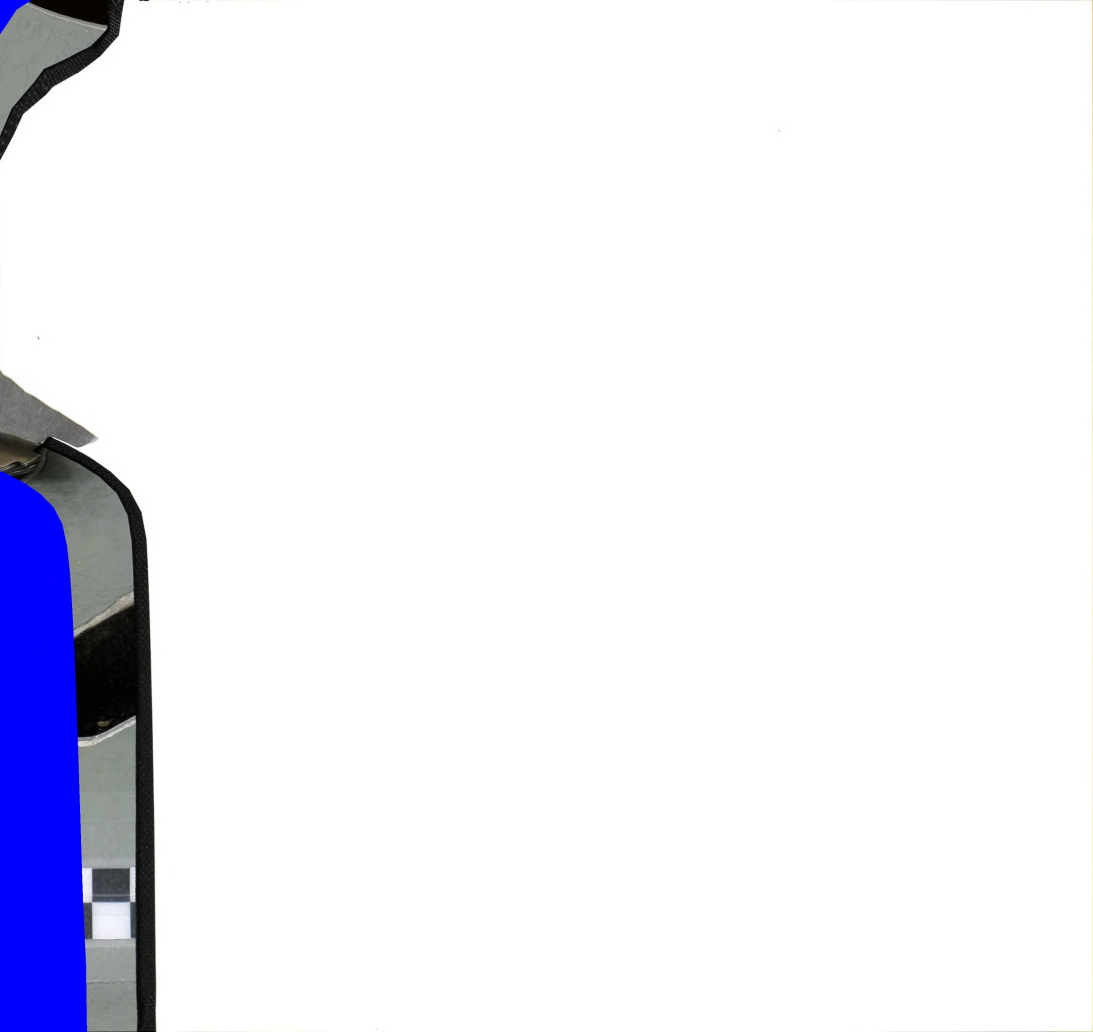
preceptors, a psychiatrist, provided an interpretation that fit the data better. He argued that the students did indeed care about the patient (and future patients) and identified with his ordeal (i.e., they exhibited Moral Regard and Empathy). According to him, the first half-hour of the discussion did not indicate a lack of interest, but instead indicated students' hope that improved technology and therapeutic techniques may have eliminated the problems the videotape raised. He pointed out that once students realized that similar problems still must be faced, and once they had "cooled off" from the emotions prompted by the videotape (and had overcome "avoidance"), they turned to a pointed discussion of the issues raised.

The above two examples illustrate how the variant of the basic functional model used in the Focal Problems course can be sensitive to indirect goals. In hindsight, a much more systematic examination of the indirect goals would have been possible. It is noteworthy, however, that the flexible and open-ended natures of unstructured observation and interviews (and thus the evaluation design as a whole) enabled the data to be analyzed and conclusions drawn post hoc. It was not necessary to plan the analyses in advance.

Evaluation Impact

The third meta evaluation question of this example is whether the basic functional model can be adapted to the medical school context and yield useful information. An important consideration in answering this question is the
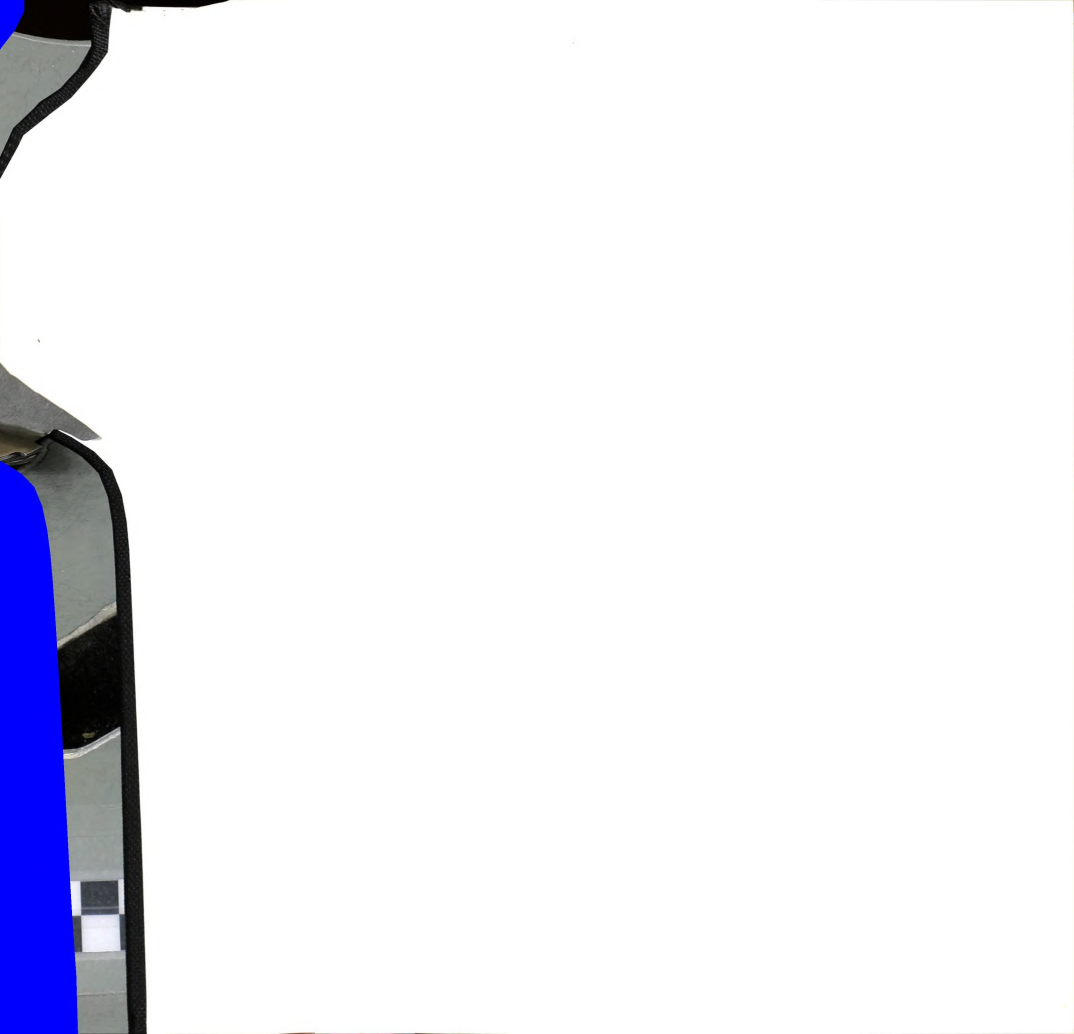
impact of the substantive evaluation findings.

As stated before, the Focal Problems course was a central proving ground in the initial stages of the "Ethics in the Core Curriculum" project; the evaluation was primarily formative. Despite the problems identified, problems potentially more damaging to the course and the project as a whole did not arise. Specifically, neither faculty nor students viewed ethics as mere window dressing. On the contrary, both groups were engaged by the issues and impressed with their importance and relevance. Moreover, the problem of subjectivism was not significant; it arose only obliquely in the claim made by some students that they should not be tested whatsoever in ethics. The evaluation of the Focal Problems course thus provided evidence early on that the project could proceed on the assumption that ethics would be well-received and not woefully misunderstood.

The evaluation also provided information useful for improving the Focal Problems course itself. Two general problems had been identified: testing and skill of preceptors in leading discussions. These findings motivated the project staff to make several changes.

Five rather than two faculty development seminars were planned for the following year and were to span the preceding term as well as the term in which the course was taught. The aim was to increase the amount of training given to the preceptors and to monitor their progress.
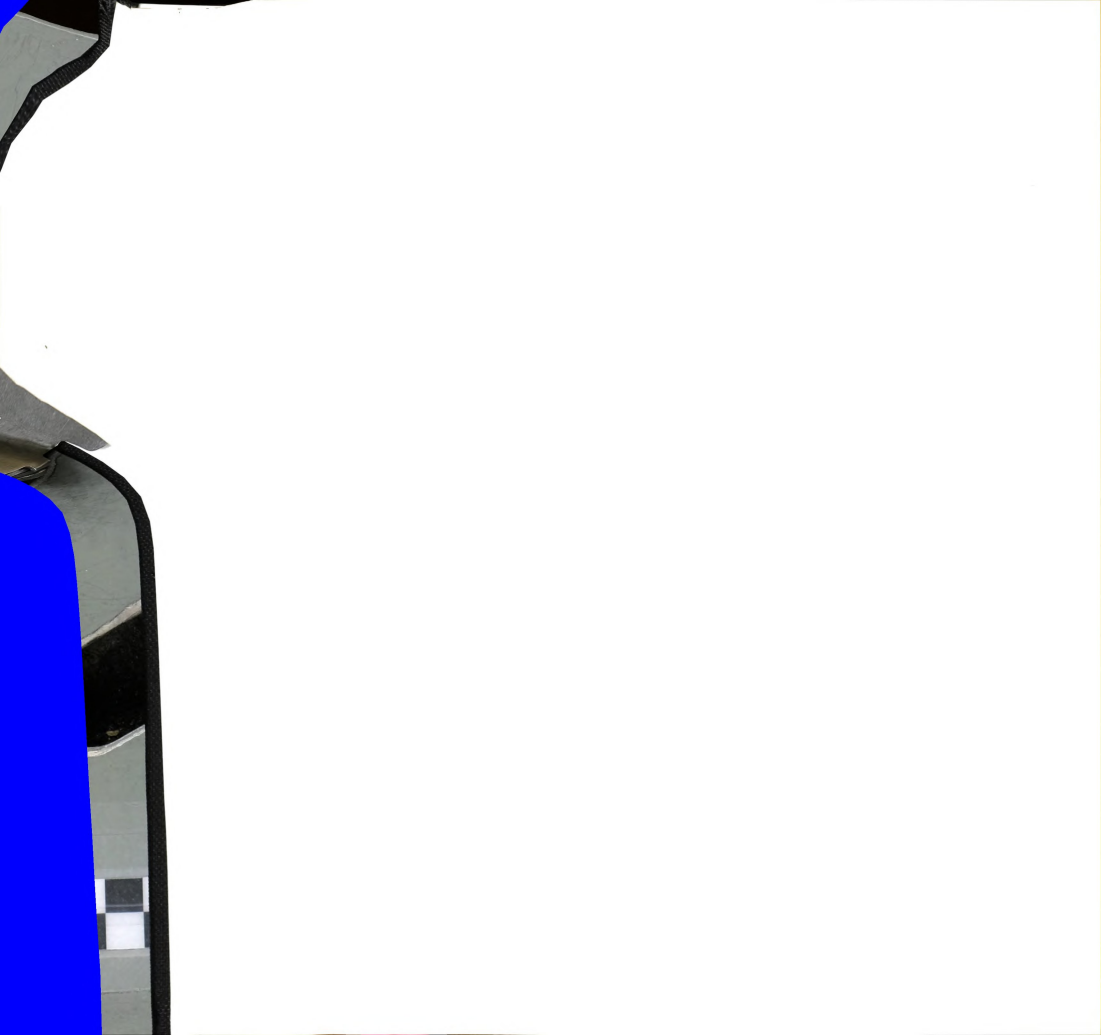
Also, the participation of preceptors who gained experience the first time the course was taught was solicited. Other changes included matching experienced preceptors with inexperienced ones, withholding the leaders' guides from students, and making some revisions in the case materials. Changes in testing methods were also made, and these will be considered in the next example.

Conclusion

Three meta evaluation questions were the focus of the Focal Problems evaluation: (1) Can fixed-response tests supplant essay tests as a measure of Knowledge and Reasoning? (2) Can the variant of the basic functional model used to evaluate the Focal Problems course detect evidence for indirect goals? and (3) Can the model be adapted to the medical school context and yield useful results? The answers to each of these questions may now be summarized.

(1) Testing exclusively with fixed-response examinations is ill-advised. Although construct valid fixed-response tests of Knowledge and Reasoning can probably be devised, other considerations argue against their use. Like Kohlbergian measures, such exams are rarely useful for providing feedback to students and thus have little pedagogical value. Also, because they lack face validity, such exams are uninterpretable to students and to faculty responsible for teaching, and can even engender hostility. Finally, the resources required to develop defensible fixed-response measures of Reasoning are significant.

Overall, the possible advantages of fixed-response tests of Reasoning for reducing fallibility are outweighed by the practical disadvantages of such tests.

(2) Unstructured observations and interviews provided evidence regarding the indirect goals. Although the evidence was somewhat thin and the analysis somewhat unsystematic, two provisional conclusions may be drawn. First, the basic functional model, in virtue of incorporating observation and interviews, can detect evidence of the indirect goals at the same time it detects evidence of the direct goals. This renders the model both functional and efficient. Second, goals like "stimulating Moral Regard", "eliciting Empathy", "enhancing Interpersonal Skills", and "promoting Courage" are (or can be) a part of medical ethics teaching, and this provides additional support for the arguments in Chapter II about the appropriateness and role of the indirect goals.

(3) The variant of the basic functional model used in the Focal Problems course produced findings regarding Knowledge, Reasoning, and Appreciation that were less positive and more uncertain than those of the ethics in nursing course. The testing results in particular were disappointing--the failure to demonstrate cognitive effects left the door open for the scientistic charge that ethics teaching and its evaluation are "soft", and that the latter consists merely in collecting "happy data".

On the other hand, the general design of the Focal

Problems evaluation generated results (in terms of its "soft" qualitative indicators) that were credible to the "Ethics in the Core Curriculum" program staff and to external consultants. The results of the evaluation (a) warranted the conclusion (in light of the difficulties experienced by preceptors) that ethics teaching was more difficult and specialized than originally believed; (b) stimulated changes in the Focal Problems course itself (described above); (c) combined with other attitudinal data collected regarding the "Ethics in the Core Curriculum" project to reduce apprehension about possible resistance to the introduction of ethics into nursing and medical education; and (d) provided the first data regarding responses to (versus predispositions toward) required ethics curricula in these contexts. In short, the results of the Focal Problems evaluation had a significant formative impact on the conduct of the "Ethics in the Core Curriculum" project. The functional evaluation approach employed in the Focal Problems course thus achieved its major substantive purpose.

Example 3: Focal Problems (1983)

Course Description

The 1983 Focal Problems course had the same basic format as the 1982 version: one hour of lecture per week in joint meetings combined with two-hour small group discussions led by preceptor pairs. The 1983 course differed from the 1982

one in terms of the previously mentioned changes in faculty development and staffing motivated by the 1982 evaluation. Changes in course materials were also motivated by the 1982 evaluation results: the senile dementia case was deleted, the Karen Quinlan case was pared back, and two new cases were added. The two new cases involved a suicidal multiple sclerosis patient and a terminally ill leukemia patient under consideration for a research protocol. These cases were less well integrated with the remainder of the curriculum than the one deleted and placed less stress on biomedical knowledge, both of which were also justified on the basis of preceptor and student opinion from the previous year.

Evaluation Purposes

The evaluation of the 1983 course continued with the substantive purpose of improvement which had been central in 1982. Regarding the Focal Problems course, it was necessary to investigate the success of the changes which had been instituted (such as increased faculty development), the change in cases, and the feasibility of using clinical faculty to grade essays (the implementation of essay tests is described below). More broadly, results from 1982 Focal Problems, coupled with evaluation results from other activities of the "Ethics in the Core Curriculum" project (well into its second year by this time), dictated that evaluation should focus on an investigation of cognitive effects. In light of the mixed results on cognitive effects

from the Focal Problems (1982) evaluation, the "Ethics in the Core Curriculum" project needed to refute the scientistic notion that ethics teaching and its evaluation are inherently "soft". Student and faculty Appreciation (or at least endorsement) seemed widespread, and the Focal Problems course still presented the best opportunity to investigate cognitive effects because it remained the most circumscribed context and involved the most intensive exposure.

Because the substantive evaluation aim was by-and-large limited to investigating cognitive effects, the Focal Problems (1983) evaluation focused more explicitly on meta evaluation issues than the previous two examples. Three meta evaluation questions were of particular concern: (1) Can Knowledge be fruitfully distinguished from Reasoning in terms of testing (where Knowledge is identified with what is measured by fixed-response tests and Reasoning is identified with essay tests)? (2) Can essay tests on ethical issues be scored with a reasonable degree of inter-rater reliability? and (3) Can a testing strategy that combines fixed-response measures of Knowledge with essay measures of Reasoning avoid the practical difficulties that attend more formal and abstract measures (such as Kohlbergian measures and the pre-post test used in the Focal Problems, 1982, evaluation)?

As in the previous examples, the discussion of Focal Problems (1983) focuses on meta evaluation: the remainder of the discussion will concentrate on answering the above three
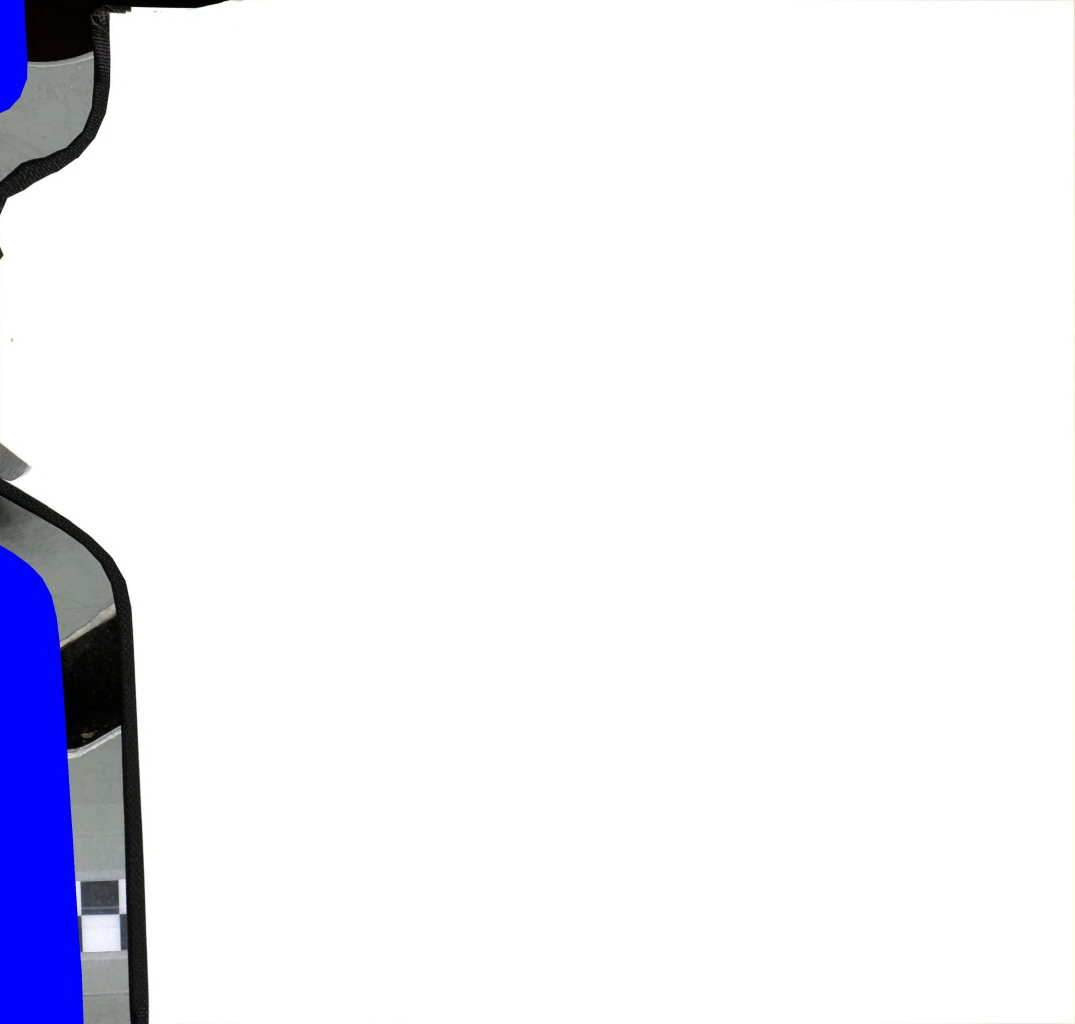
questions. Substantive evaluation findings are discussed when appropriate.

Evaluation Design

Scientistic evaluation tends to employ one-shot quantitative studies, feigning ignorance of (or disparaging) relevant (often "qualitative") background knowledge. A functional approach holds that individual evaluations should not "stand alone", i.e., relevant knowledge (quantitative and qualitative) should be put to use in devising methods.

The design of the Focal Problems (1983) evaluation made liberal use of pre-existing knowledge. In the 1983 version of Focal Problems, exams that combined fixed-response items with short essays replaced the 1982 strategy of exclusive reliance on fixed-response tests. In addition to the Focal Problems experience from the previous year and the successful use of essay examinations in the nursing ethics evaluation, two additional sources of data suggested that essay tests might be more workable than previously believed. First, the difficulty in training non-philosophers to grade essays may have been overestimated. A structured list of criteria was devised to aid a clinician in grading ethics case write-ups in 3rd year clerkships, and it met with reasonable success. Second, student resistance to writing may also have been overestimated. Students gave the following as their first choices of evaluation on the 1982 Focal Problems end-of-term course evaluations: 28% for fixed-response; 23% for short answer; 12% for in-class

essays; and 35% for short take-home papers (3-5 pages).

The distinction between Knowledge and Reasoning crystallized in terms of instruments. In the Ethics in Nursing and Focal Problems (1982) evaluations, single tests (essay and fixed-response respectively) were used to measure both Knowledge and Reasoning. In the Focal Problems (1983) evaluation, Knowledge of content was identified with what would be measured by fixed-response tests and Reasoning with what would be measured by short essays regarding novel medical-ethical problems. It was reasoned that fixed-response tests could adequately measure recall and recognition of reading and lecture content, but were unwieldy and beset with practical difficulties in connection with Reasoning. The findings from the Ethics in Nursing evaluation suggested that essay tests could be used to measure Reasoning.

Employing this strategy, two measures were developed, test forms A and B. Each form had 12 objective items based on course content (readings and lectures) and a short essay on a medical-ethical problem not discussed in the course. The essays required students to indicate whether they agreed with the position taken in the case description provided and, independent of this, whether the support provided for the position was satisfactory. Each form of the exam required approximately 35 minutes, and was administered in the first small group meetings and again as a portion of the final examination for the course.
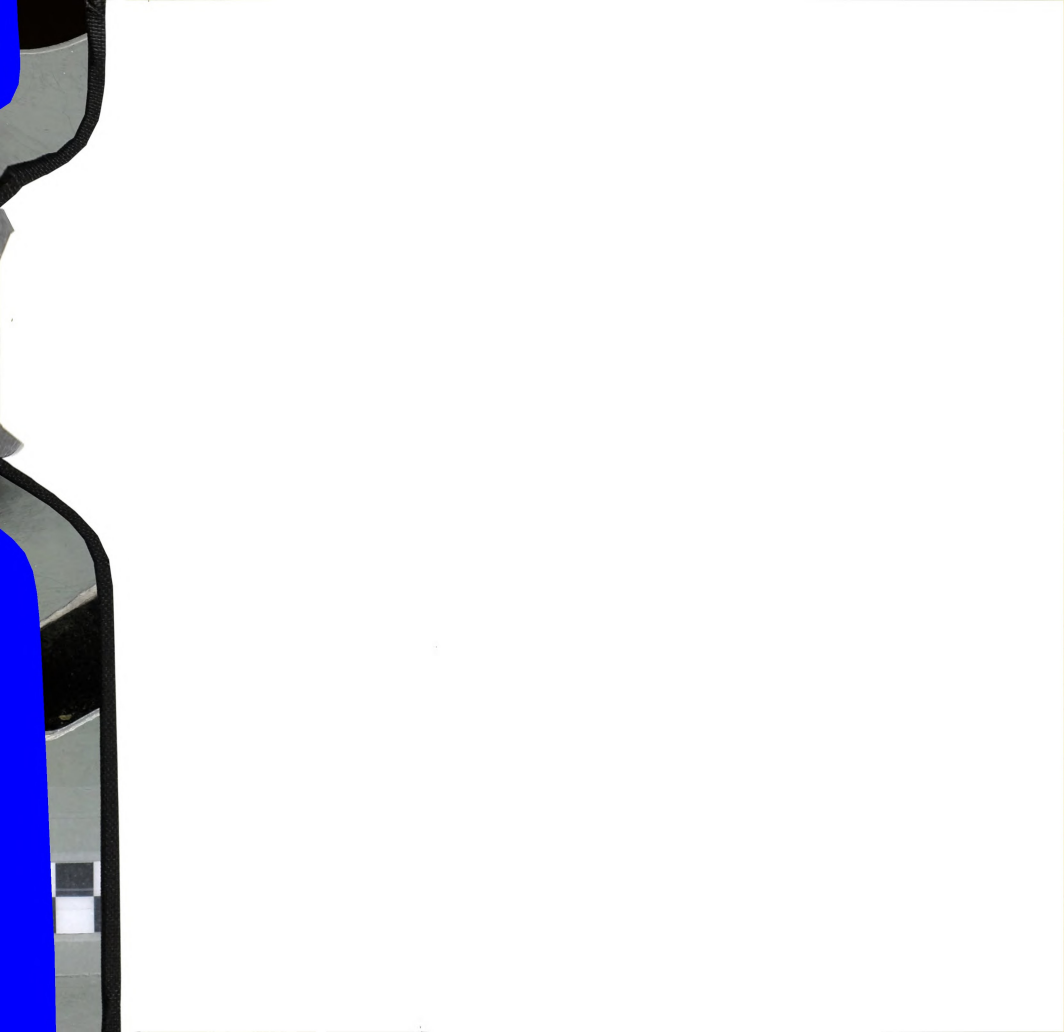
A more sophisticated design was employed than the X O X type of the previous two examples. The students were divided into two groups. The first group consisted of 28 students who met in small groups on Wednesdays; the second group consisted of 32 students who met in small groups on Thursdays. The group which took form A of the test as a pre-test took form B as a post-test and vice versa. This crossed design improved confidence about conclusions over the X O X design by controlling for testing effects and providing two replications at once. Although an X O X design would have been adequate, the greater reduction in fallibility afforded by the more sophisticated design was consistent with a functional approach because the greater reduction in fallibility could be purchased at an acceptable cost: two measures had to be devised rather than one, and the data analysis was somewhat more demanding.

This investigation of cognitive effects addressed the substantive question of the impact of the Focal Problems course regarding learning. It was also bound up with the investigation of several ancillary meta evaluation questions. First, an examination of correlations between scores on the fixed-response and essay portions of the exams would shed light on the practical difference between Knowledge and Reasoning. Second, study of the correlations (or lack thereof) between the positions taken on the essays and the ratings received would constitute evidence for (or against) the claim that essay grading is biased. Third, if
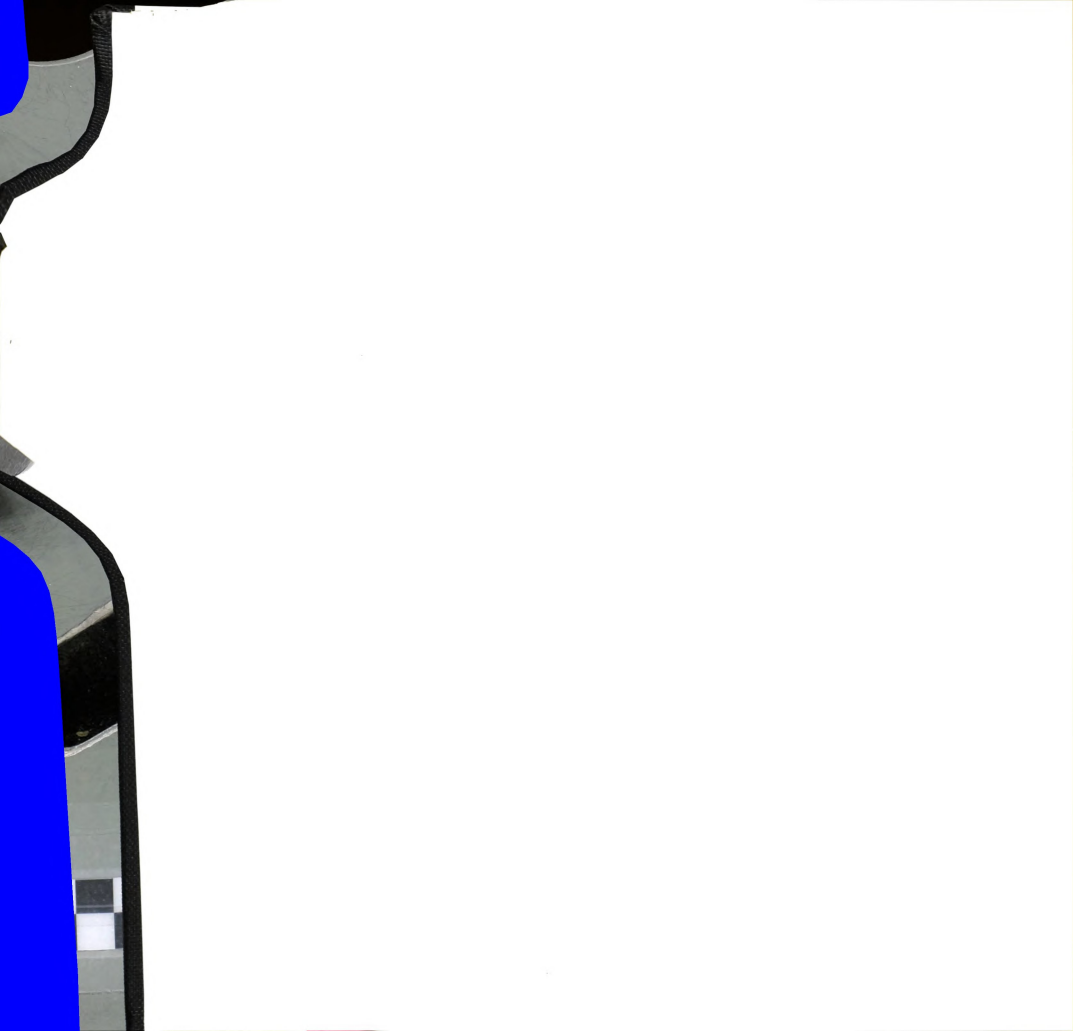
successful, the strategy of combining fixed response and essay tests would constitute a face valid, practicable alternative to other means of evaluating cognitive performance such as those based on Kohlberg's theory.

The design of the 1983 Focal Problems evaluation emphasized cognitive effects but was not confined to these. Though observation of the small groups and formal interviews of preceptors were not employed as they had been in 1982, the progress of the preceptors was monitored by informal discussions within the expanded faculty development sessions. Again, evaluations should not stand alone. Foregoing direct observations and formal interviews was justified in light of the resources required for the more elaborate effort of 1982; the reduced likelihood that much new would be learned as a result of the similar findings from the Ethics in Nursing and Focal Problems (1982) evaluations; and the changed make-up of the 1983 cohort of preceptors such that each group had an experienced preceptor, one from the staff of the MHP, or both.

The 1983 design included student evaluations for the same purposes as before. The design also provided for an investigation of the inter-rater reliability of essay grading, an issue which was introduced with the change in the testing format.

Figure 6 summarizes the relationships between data collection methods and constructs measured. (Because of the emphasis on meta evaluation in the design of the evaluation,
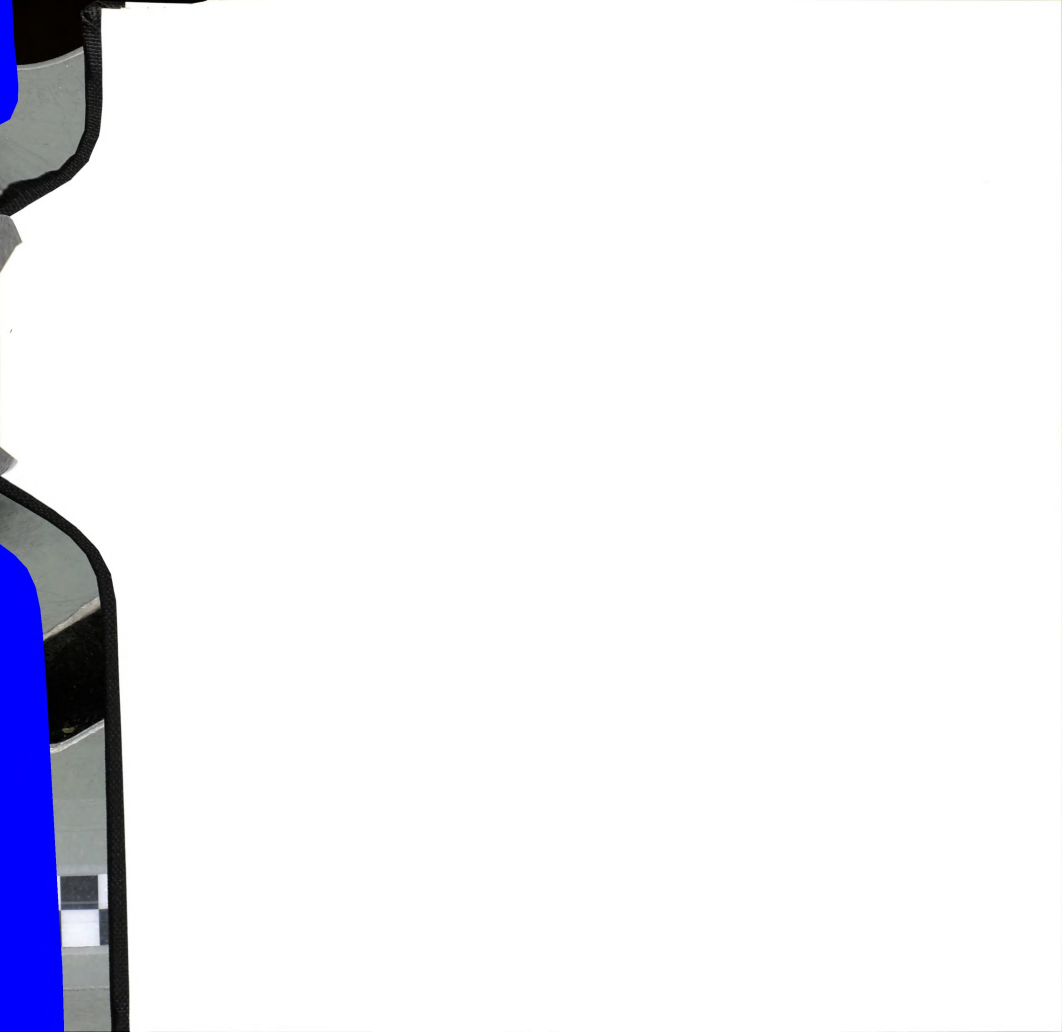
the figure is less informative as a guide to the discussion of this example than the previous figures were. It is included for consistency and as a characterization of the substantive evaluation.)

	Knowledge	Reasoning	Attitudes
fixed-response tests	x		
essay tests		x	
evaluation forms			x
interviews (informal)	x	x	x

Figure 6. The Relationships Among Data Collection Techniques and Constructs for Focal Problems (1983)

Evaluation Results

Following the pattern of the two previous examples, the report of results is keyed to meta evaluation questions. To reiterate, the Focal Problems (1983) study involved three such questions: (1) Can Knowledge be fruitfully distinguished from Reasoning in terms of testing (where Knowledge is identified with what is measured by fixed-response tests and Reasoning is identified with essay tests)? (2) Can essay tests on ethical issues be scored with a reasonable degree of inter-rater reliability? and (3) Can a testing strategy that combines fixed-response measures of Knowledge with essay measures of Reasoning avoid the practical difficulties that attend more formal and abstract measures (such as Kohlbergian measures and the pre-post test used in the Focal Problems, 1982, evaluation)?

The subsequent "evaluation impact" and "conclusion" sections also follows the previous pattern. The "evaluation impact" section argues that the variant of the basic functional model used in the Focal Problems (1983) established worthwhile and credible conclusions; the concluding section summarizes the major meta evaluation conclusions of this example.

Analysis of Pre-post Testing. The crossed design and the nature of the data (especially the essay ratings) suggested the Mann-Whitney U as the appropriate means by which to analyze the pre- post-test data. Essays were graded using the same blind procedure used in the nursing example. Four comparisons were made; results are reported in Table 2.

Table 2. Comparison of Two Groups on Fixed-Response and Essay Tests

	Pre-test*			Post-test*		
	Test Form**	Mean Score	S.D.	Test Form**	Mean Score	S.D.
Group 1	Af	7.82	1.93	Bf	8.93	1.36
(n=28)	Ae	2.13	.88	Be	3.31	.58
Group 2	Bf	7.50	1.98	Af	8.72	1.59
(n=32)	Be	2.66	.87	Ae	2.80	.89

*Pre-post comparisons for Af, Bf, and Be were significant at $p < .002$; for Ae, $p < .012$.

**A = test A; B = test B; f = fixed response test, possible range 0-12; e = essay test, possible range 0-4.

In contrast to the 1982 Focal Problems testing, these results provided evidence based on testing that the course

did produce cognitive gains. Although this was not buttressed by observational evidence, preceptors provided informal testimony within the faculty development sessions and it was reasonable to assume that observations of the groups would have provided the same kind of evidence as they had in the 1982 evaluation. The results of the pre-post testing, then, provided rather solid evidence for the substantive conclusion that cognitive learning had occurred, especially when these results were combined with the previous findings from Focal Problems and the ethics in nursing course.

These findings also helped in the inference from positive student evaluations of the course to Appreciation. Unlike the 1982 Focal Problems course, students' favorable evaluations of the course could not plausibly be attributed to merely wanting an easy course. Demonstrated cognitive effects (coupled with evidence from multiple indicators in relevantly similar circumstances) rendered this explanation untenable.

Answers to the three meta evaluation questions of focus in this example were obtained by further analysis of the testing data.

Distinguishing Knowledge and Reasoning in Terms of Testing. Results on the Mann-Whitney U tests indicate that the tests employed were sensitive to Knowledge and Reasoning and therefore are valid measures of these two goals. The apparent circularity of this argument (namely, the

instruments measured something, therefore they measured what was of interest) may be mitigated by making several observations. First, although such an analysis was not performed, one could expect the same differences to be apparent in the essays that were documented in the ethics in nursing study. This is supported by the fact that, when debriefed, the grader of the pre- post essays in Focal Problems reported that he detected differences in the essays in terms things such as listing alternatives and responding to objections. Second, observational evidence from the nursing study and the 1982 Focal Problems course provides independent evidence that learning of the desired kind occurs in similar courses. Third, the fixed-response exams were judged to be face valid measures of Knowledge and the essay exams were judged to be face valid measures of Reasoning by the MHP staff, as well as by preceptors and students. This latter characteristic as well as sensitivity to effects are possessed neither by the type of instrument used in the 1982 Focal Problems study nor by Kohlbergian-type measures.

Certain findings from the testing add to the appeal of the distinction between Knowledge and Reasoning. If these are indeed distinct, one would expect correlations between tests of them to be low. Goodman's and Kruskal's gamma was used to investigate the relationship between the fixed-response and essay portions of the two test forms, A and B. Only one of these was significant, namely, the

portions of B as a pre-test (gamma = .348, p < .013). Further support for the distinction derives from considerations of face validity mentioned above, and the fact that the ratings on the essays did not correlate significantly with the positions defended in the essays (tau = .215 for the correlation between position and rating on the form A essay and .132 on the form B essay). This supports the contention that the essays measured (and were rated on the basis of) Reasoning (or quality of argument) and not on substantive claims. In connection with this, the essays involved cases not pursued in Focal Problems. Students were required to determine for themselves what content from Focal Problems was relevant and how it might be applied.

Inter-Rater Reliability. The second meta evaluation question in the Focal Problems (1983) study was inter-rater reliability. A significant shortcoming of essay tests is lack of consistency among graders. Many believe (scientistic evaluators among them) that this shortcoming is especially acute in ethics because there are no "right answers".

The 1983 cohort of preceptors were required to grade 2 sets of essays. They were provided with general grading criteria and broad outlines of anticipated arguments. Four pairs followed the instruction to grade independently, and inter-rater agreement (calculated using the Spearman rank-order correlation coefficient) was respectable. A
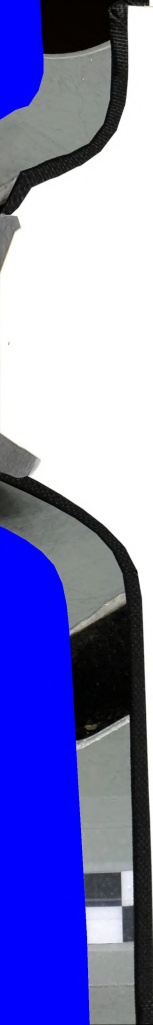
z-transformation (using r to approximate r_s) yielded a weighted average of .80; the individual coefficients are reported in Table 3.

Table 3. Reliabilities of Preceptor-Pairs' Essay Grading

Preceptor pair	1rst essay	2nd essay**
A*	$r_s = .68(n = 12)$	$r_s = .70(n = 11)$
B	$r_s = .82(n = 12)$	$r_s = .73(n = 12)$
C*	$r_s = .83(n = 9)$	$r_s = .89(n = 9)$
D*	$r_s = .90(n = 10)$	$r_s = .76(n = 10)$

*One member of the pair was a philosopher

**Instructions were changed for the second essays. Both were graded on a 0-10 scale, but 7 was explicitly defined as minimally adequate on the second set to reflect the customary 70% pass level in the college. This had the effect of restricting the practical range of scores on the second set as compared to the first, accounting for the slight decrease in overall agreement.

In addition to these results which testify to the feasibility of using non-philosophers to grade essays on ethical problems, the preceptors also claimed that the essays were useful for providing feedback. Students confirmed this on the evaluation forms and expressed much less dissatisfaction with testing in general than the 1982 cohort. Furthermore, few students claimed that the grading was subjective or biased toward the favored views of the graders.

Usefulness of Measures of Knowledge and Reasoning. The third meta evaluation question was whether using fixed-response and essay tests to measure Knowledge and Reasoning respectively can avoid the practical problems associated with more abstract measures.
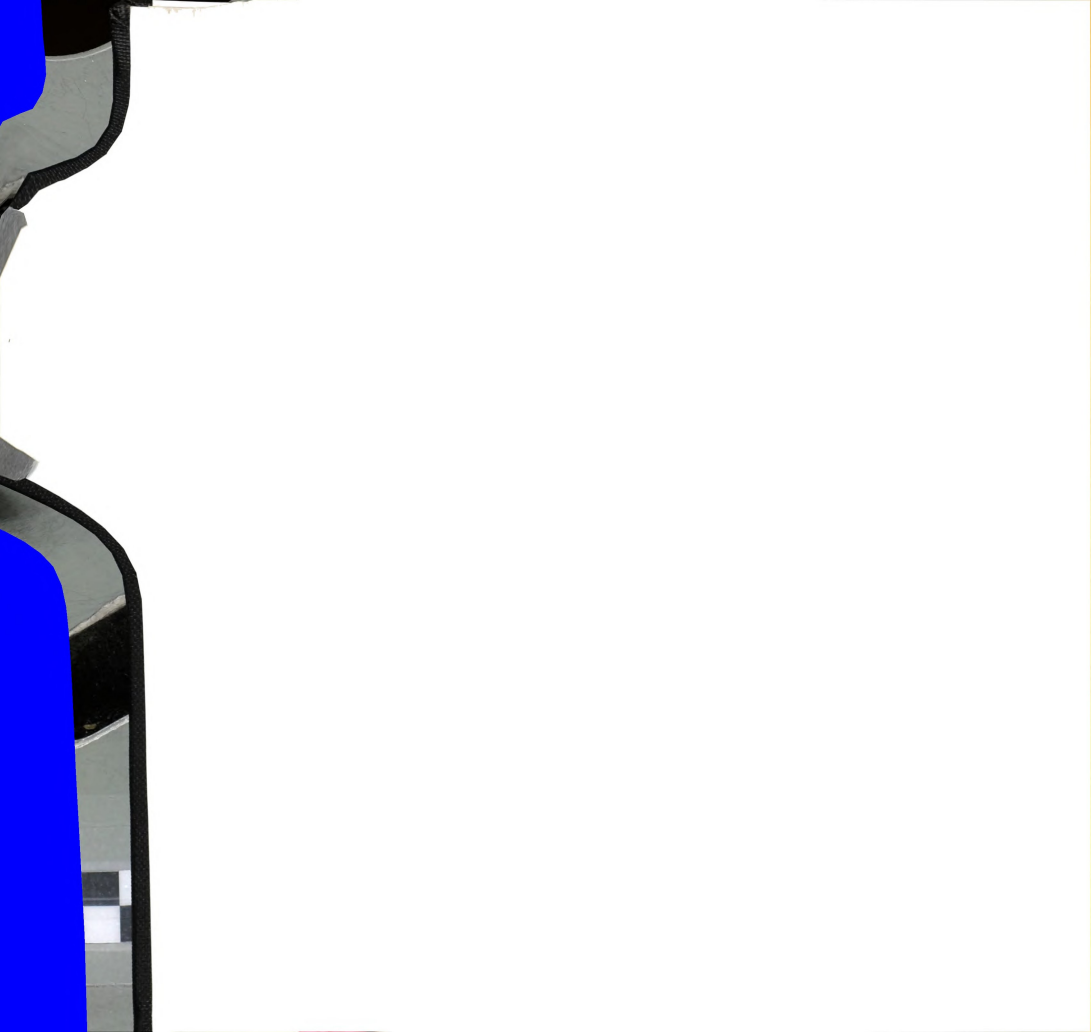
In the Focal Problems (1983) evaluation, keying fixed-response tests to Knowledge and essay tests to Reasoning provided formative as well as summative information. Table 4 contains 3 items from the fixed-response portion of form A of the pre- post-test, accompanied with correct responses (indicated by *) and pre-post difficulties (percentage of students responding incorrectly).

Table 4. Pre-Post Difficulties of Selected Items

Item	Difficulties Pre(n = 28) Post(n = 32)	
1. The right of a research subject to informed consent requires	9	0
a) informing subjects of all alternatives		
b) informing subjects of the risks and benefits of the research procedures		
c) informing subjects that they may withdraw at any time		
d)* all of the above		
2. The right to autonomy (self-determination) is included in the constitutional right to privacy	47	16
a)* true		
b) false		
3. The right to life implies that life saving medical treatment may never be withheld	53	72
a) true		
b)* false		

These three items may be interpreted in the following way. Item 1 is indicative of progress regarding the knowledge in question. Because the item was so easy to begin with (only 9% answered it incorrectly), however, it is
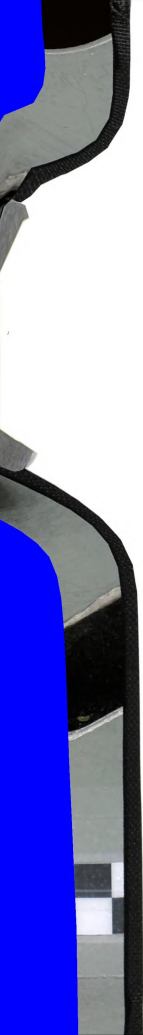
more useful for diagnosis than assessment of learning. That is, students apparently had the desired knowledge entering the course, and this would warrant the decision to give little attention to the issue in future offerings of the course. Item 2 shows a substantial gain in knowledge about the relationship between law and morality, and is useful as a straightforward gauge of knowledge acquisition. Item 3 shows a loss, and a reasonable conclusion is that the instruction confused the students. For instance, perhaps the students' respect for the rights of patients increased but was not accompanied by a sufficiently sophisticated understanding of rights (e.g., the right to life does not entail the right to an artificial heart).

Essay exams augment the fixed response tests and may be used for similar purposes (i.e., diagnosis and assessment of gains) with respect to Reasoning. Although interpretation of essay test performance is rarely as straightforward as interpretation of fixed-response test performance, essays as pre-tests provide an indication of whether students can put together coherent positions and, accordingly, help determine how much energy should be put in this direction. They also may be used to trace progress, as was done in the nursing and Focal Problems 1983 evaluations.

Two points made earlier in different contexts may be expanded and applied to the question of the usefulness of the combined fixed-response/essay testing strategy. First, fixed-response and essay exams of the type used in the Focal

Problems (1983) course are face valid measures of Knowledge
and Reasoning respectively. Because the measures had face
validity, they were useful for feedback and were
interpretable to the "Ethics in the Core Curriculum" project
staff, the preceptors, and the students. Lack of face
validity is a major shortcoming of Kohlbergian-type measures
and the abstract kind of fixed-response test used in the
Focal Problems (1982) evaluation. Second, and with respect
to essay tests in particular, essay tests on ethical issues
can be scored with an acceptable degree of inter-rater
reliability and do not automatically prompt charges of
subjectivity and bias from students. Furthermore, essay
tests have a distinct advantage over fixed-response measures
of Reasoning in terms of the time, effort, and expertise
required for development.

Evaluation Impact

The 1983 evaluation results demonstrated the value of
the changes prompted by the 1982 evaluation results. In
addition to the solid findings regarding cognitive effects
(and the related meta evaluation findings), students'
evaluations of preceptors' performance improved from the
previous year. The higher ratings given to the lectures in
general and the small group involving a philosopher
preceptor in 1982 diminished. In 1983, 3 of the 6 groups
had philosopher preceptors and differences between these
groups and the others were negligible, as were the ratings
of lectures versus small group discussion. In addition to

data from student evaluations, the preceptors reported in the faculty development sessions that things went well and those with experience reported greater confidence the second go around. Notably absent were reports of discussions stalling out, not knowing the right "moves", and so on.

The results from the 1983 evaluation of Focal Problems (like the results from the Ethics in Nursing course) were interpreted to mean that a viable course had been developed. Like both of the previous examples, the functional evaluation approach again produced credible results. The relative success of the course and the strategy of distinguishing and devising separate tests for Knowledge (content) and Reasoning (problem-solving) were disseminated in the form of journal articles (Howe et al., 1984; Howe and Jones, 1984). The written materials for the course have been widely distributed, and the evaluation findings have been used in arguments presented to the college's curriculum committee.

Conclusion

This example emphasized the study of cognitive effects of medical ethics teaching, and both substantive and meta issues were addressed.

The substantive findings of gains in terms of the cognitive goals of medical ethics teaching corroborate the strong findings of the Ethics in Nursing example (based on testing and observation) and the more tenuous findings from the Focal Problems (1982) example (based on observation and

interviews but contradicted by testing). The pre-post testing gains also reinforce a point made in the nursing example: the gains count against the subjectivist view of ethics implicit in the scientistic fact-value distinction. Students gained on something, and doubt about whether these gains were cognitive ones is silly in light of the evidence which has been adduced from the two previous examples in addition to this one. Moreover, the inter-rater agreement exhibited by the preceptors is prima facie evidence for the existence of agreed upon (i.e., objective) standards for judging the quality of ethical arguments that (contra Nobel, 1982) extend beyond the interests of a specialized group of philosophers.

The three meta evaluation questions of this example were: (1) whether the constructs Knowledge and Reasoning could be fruitfully distinguished in a way that identifies Knowledge with fixed-response tests that measure course content and identifies Reasoning with essay tests that measure skills; (2) whether essay tests on ethical issues can be scored with a reasonable degree of inter-rater reliability; and (3) whether the combined fixed-response/ essay testing strategy avoids the practical problems that attend abstract measures.

The combined fixed-response/essay testing strategy did well on each question.

(1) The low correlations between the fixed-response and essay tests support the contention that Knowledge and
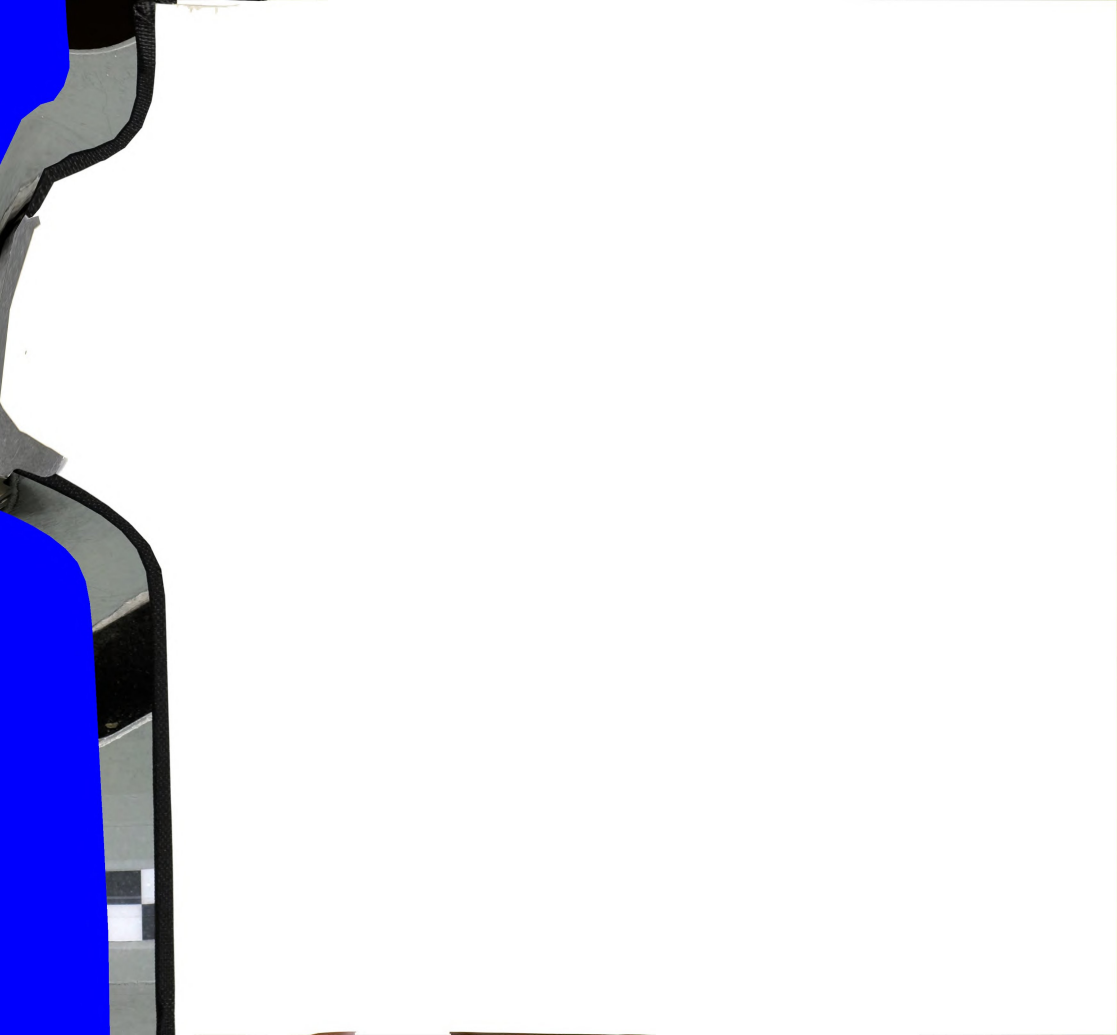
Reasoning are separable constructs that can be independently measured. Although one might argue that the correlations are "in reality" higher (e.g., because of the restricted range of the scales), the evidence militates against the conclusion that Knowledge and Reasoning amount to the same thing. The theoretical conclusion is that the distinction made between Knowledge and Reasoning in Chapter II is warranted; the practical conclusion is that fixed-response tests of Knowledge and essay tests of Reasoning are not redundant, and thus it is useful to employ both kinds of tests.

(2) The finding that the inter-rater reliability of the essay test scoring was reasonably high (.80) supports the feasibility of employing essay tests of ethical reasoning. This finding helps remove one of the most serious obstacles to the use of essay tests, clearing the way for measures that are more manageable and face valid than fixed-response measures of Reasoning. The observed inter-rater agreement also helps remove the obstacle posed by the scientistic fact-value distinction.

(3) The combined fixed-response/essay testing strategy avoids the practical problems that plague the use of abstract measures of the cognitive goals of ethics teaching. The fixed-response/essay strategy may be easily keyed to the particular content and skills objectives of individual courses, rendering such tests sensitive to specific (though by no means trivial) cognitive effects. Related to this,
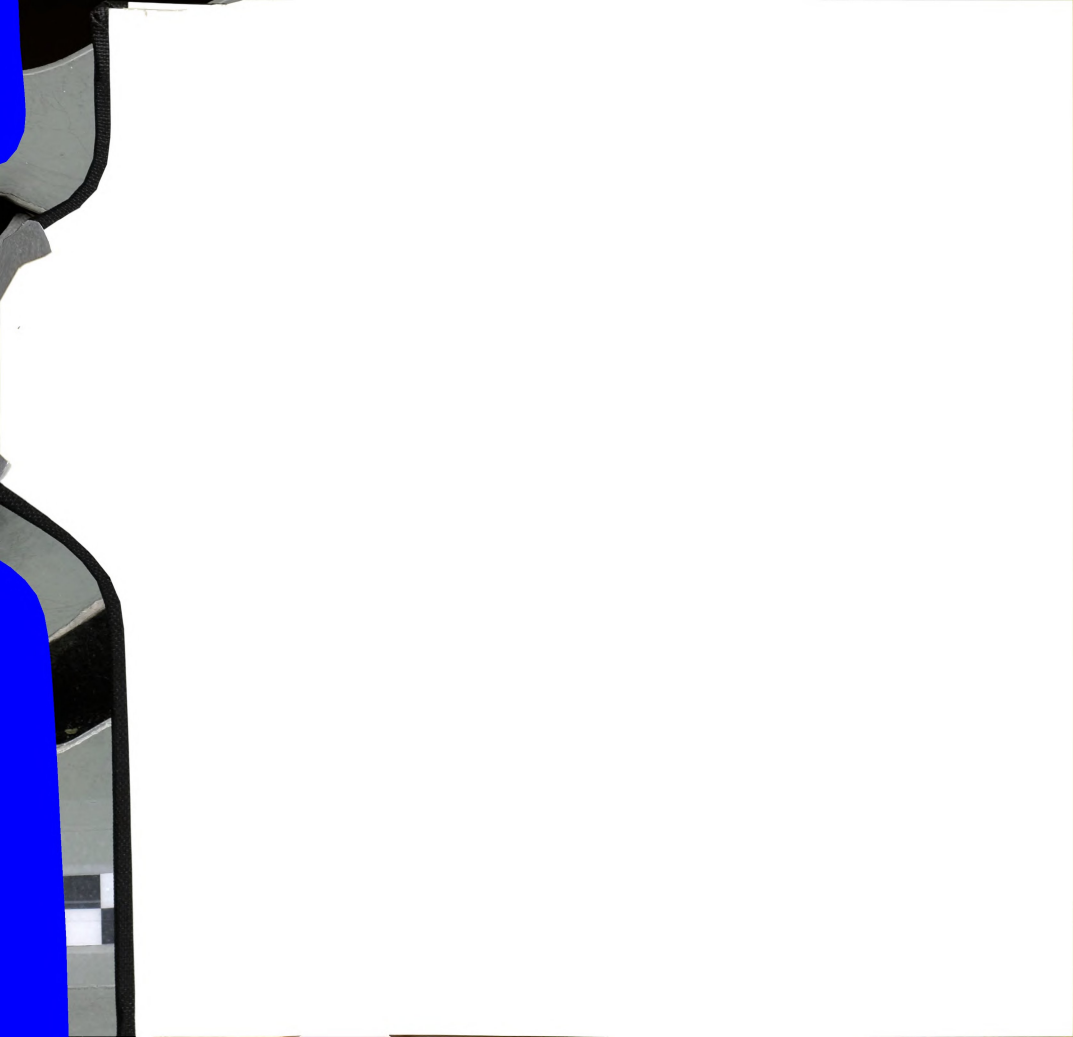
such tests are face valid and are thus interpretable to
instructors and students and are useful for feedback on
student performance. Finally, fixed-response and essay
tests on the model of the ones employed in the Focal
Problems example are easily developed and are flexible to
changes in curricular content.

Conclusion

The stated purpose of this Chapter was to show that a
functional evaluation approach, grounded in the Wilson +1
goals, can produce credible empirical results that achieve
the customary evaluation purposes of improving practice and
warranting judgments about success. To accomplish this
purpose, a basic functional model for the evaluation of
medical ethics courses was described and then variants of
this model were used to explicate three concrete course
evaluations.

Each concrete example was cast in terms of two major
themes: meta evaluation and impact. The discussions of meta
evaluation issues pertained to the credibility (in the
sense of rational warrant) of the conclusions resulting from
the variants of the basic functional model used in each
evaluation. The discussions of evaluation impact pertained
to the variants' capacities to produce results that
could influence the practice of medical ethics teaching
(i.e., the discussions of evaluation impact pertained
to credibility in a psychological, or rhetorical,
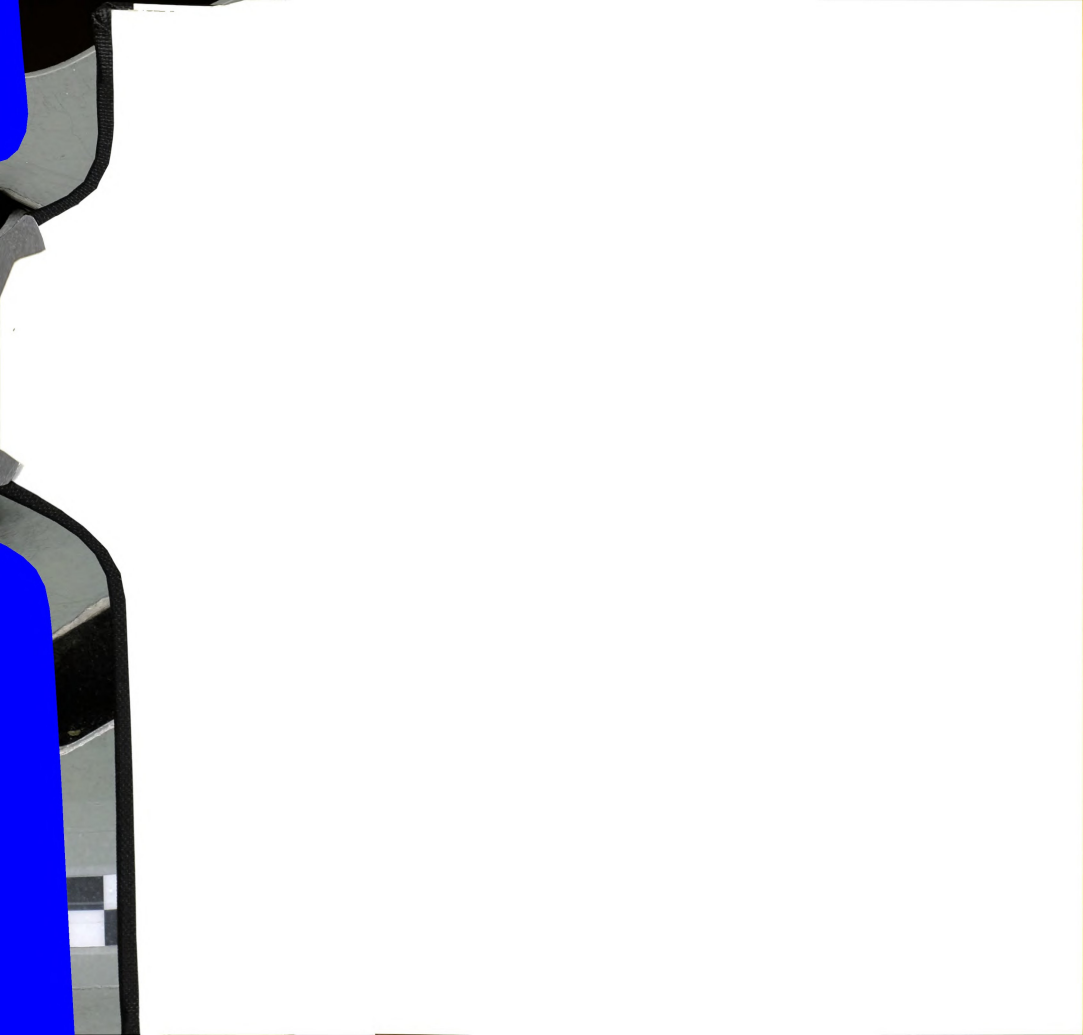
sense).

The Ethics in Nursing example provided evidence that essay tests, direct observation, and student evaluation forms can be combined to measure Knowledge, Reasoning, and Appreciation. The multiple indicators established the credibility (rational warrant) of conclusions about the measurability of these three constructs, and the impact of the findings--their influence on the instructors and their disseminability--established persuasiveness. The rational warrant and persuasiveness taken together entail that the model employed was effective. The Ethics in Nursing course was judged successful, and continues to be taught in the same way that it was in 1980 when the evaluation was performed.

The Focal Problems (1982) example showed that the basic functional model could be adapted to the medical school context and to the special multiple discussion group Focal Problems format, and that the constructs Knowledge, Reasoning, and Appreciation could be extended to faculty. In addition, the example provided suggestive evidence that direct observation (an indicator included in the basic functional model) can be sensitive to the indirect goals of medical ethics teaching. Because the fixed-response pre-post test yielded negative results, the Focal Problems (1982) evaluation did not substantiate cognitive effects. However, other indicators--observations and interviews-- helped render implausible the conclusion that students
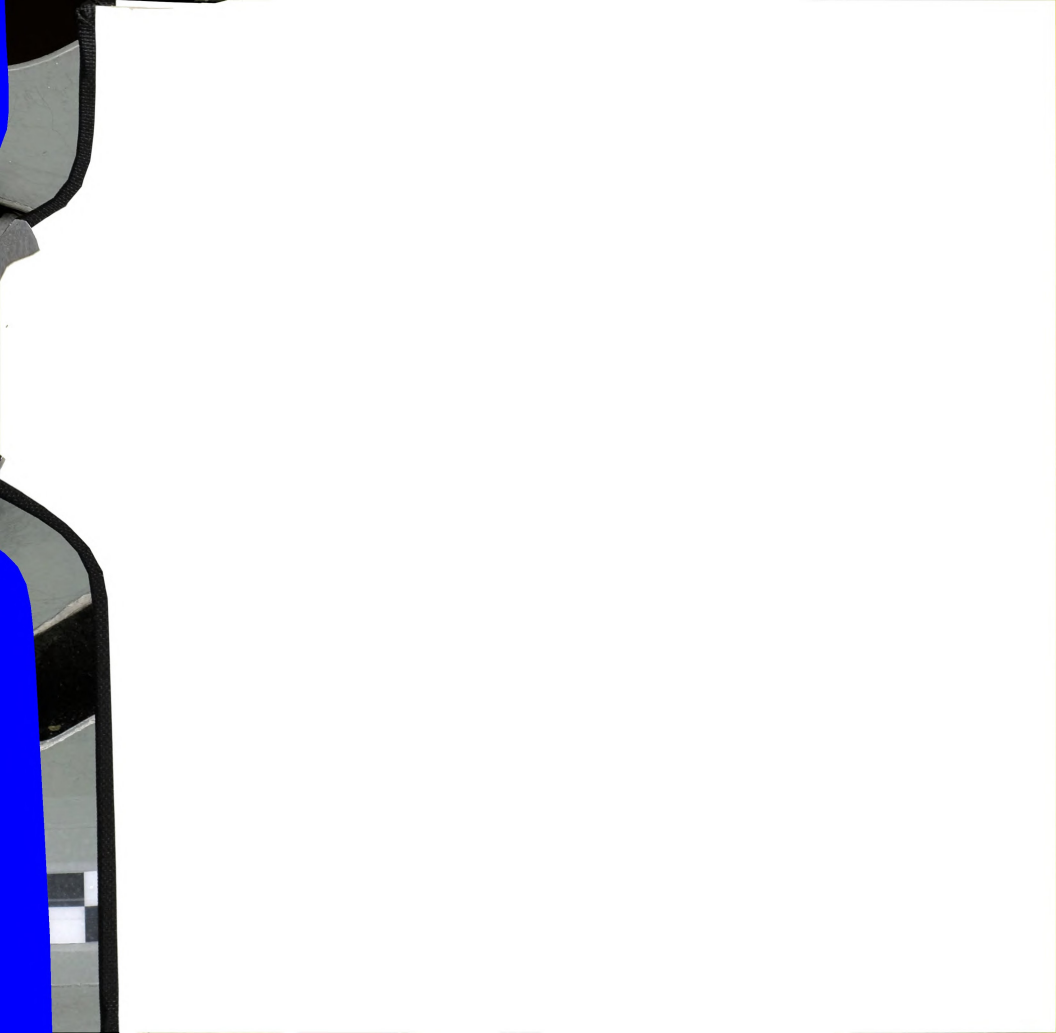
actually lost in Knowledge and Reasoning. Moreover, the unwelcomed results of the quantitative analysis of the testing did not affect the overall credibility of the evaluation. The Focal Problems (1982) evaluation identified problems with the course and prompted constructive changes.

The Focal Problems (1983) evaluation demonstrated that Knowledge and Reasoning can be fruitfully distinguished and can be measured respectively with locally constructed fixed-response and essay tests. Such a testing strategy overcomes the practical problems that beset Kohlbergian-type and other abstract measures. Like the Ethics in Nursing evaluation, the Focal Problems (1983) evaluation demonstrated that medical ethics teaching can yield cognitive gains among students. Also like the Ethics in Nursing example, the results of the Focal Problems (1983) evaluation were received as credible and as warranting the belief that a successful course (one not in need of revisions) had been developed.

The three evaluations establish that a functional evaluation approach, grounded in the Wilson+1 goals, can be effectively employed to evaluate medical ethics teaching. The examples can also be used to directly criticize the scientistic alternative. In this chapter, for instance, pre-post gains on cognitive tests and inter-rater reliability of essay ratings were used as evidence against the scientistic fact-value distinction, which implies that cognitive standards in ethics do not exist. This argument,

and others that can be based on the three concrete evaluations, will be elaborated in the next and concluding chapter.
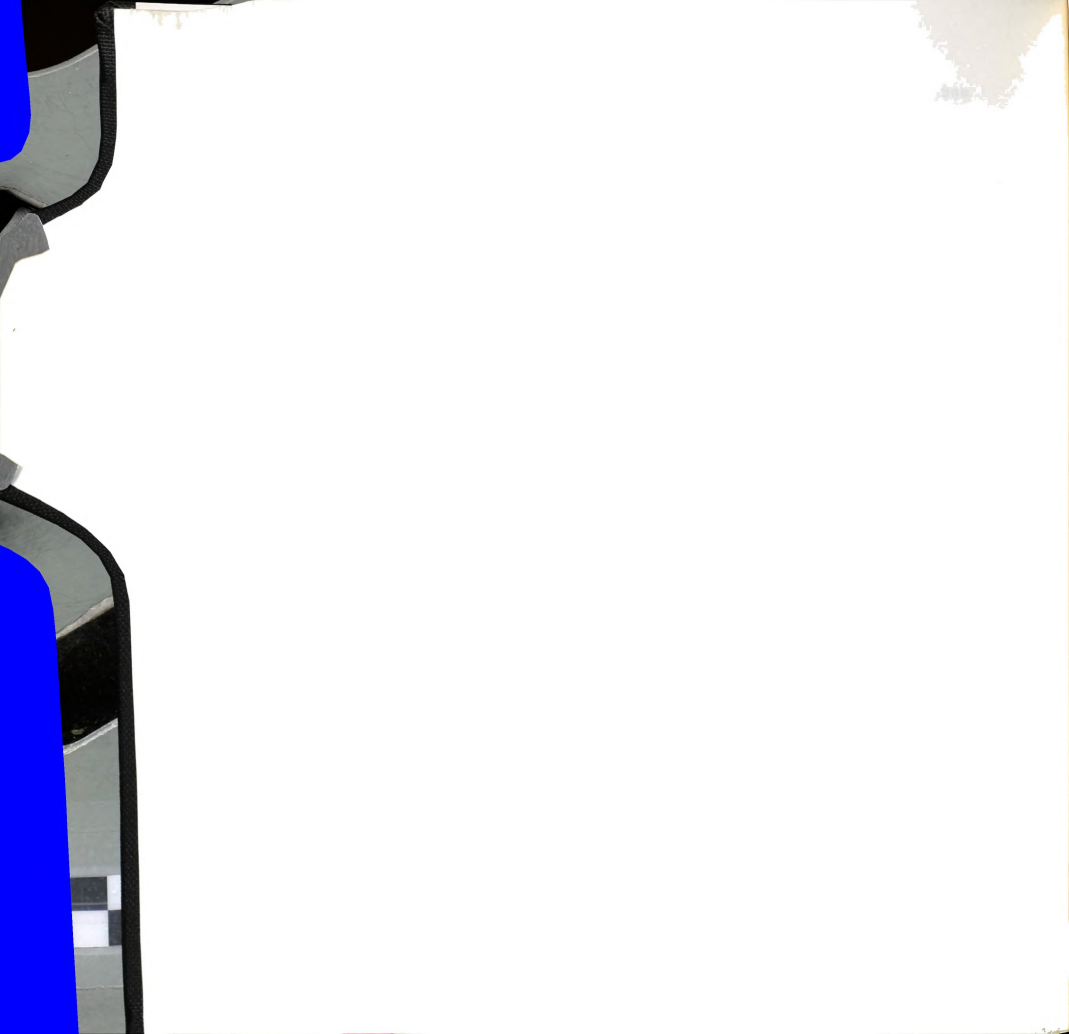
NOTES

1. Survey results from the "Ethics in the Core Curriculum" project (described in the Focal Problems, 1982, example) indicate that a majority of students and faculty believe that self-standing courses in ethics perform such a ground laying function.

2. National Endowment for the Humanities grant number EP-32926-78-1231.

3. Michigan State University's College of Human Medicine has two pre-clinical curricular options. "Track I" consists largely of lectures in the basic sciences, augmented by one three course sequence of problem-based "Focal Problems" courses. The "Track II" curricular option is composed almost exclusively of "Focal Problems" courses.

4. National Endowment for the Humanities grant number ED-20020-81-0509.
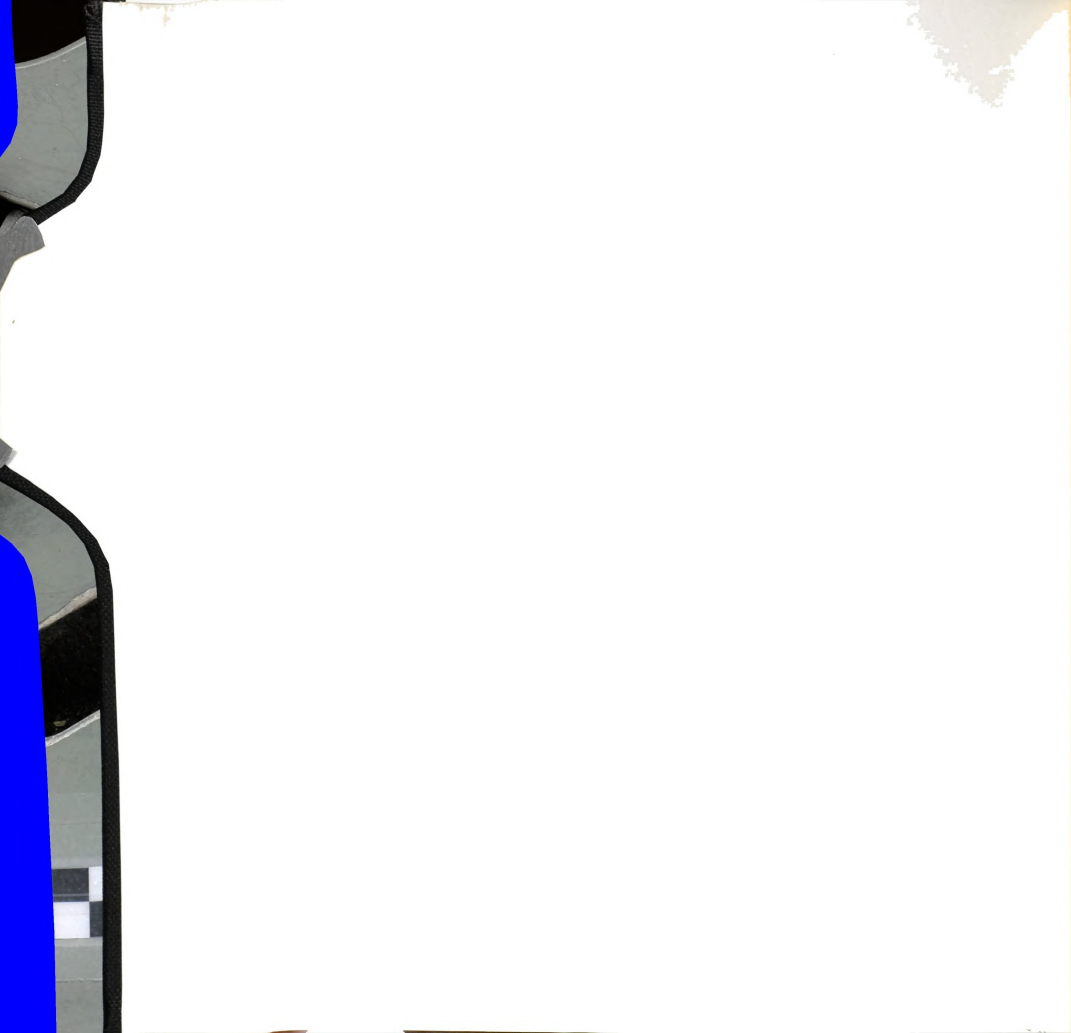
CHAPTER V

BUTTRESSING AND EXTENDING

THE FINDINGS AND ARGUMENTS

This chapter is divided into three sections. The first relates the practical examples from Chapter IV to the more theoretical and abstract themes of Chapters II and III. Whereas Chapter IV illustrated how theory is converted to and exemplified in practice, the aim in this chapter is to illustrate how practice (i.e., the three evaluations) supports theory. The discussion in the first section rounds out the central task of this dissertation: developing a defensible means of evaluating medical ethics teaching that meets the legitimate demands of both evaluation researchers and medical ethics instructors.

The second section suggests two ways of extending the evaluation model employed in Chapter IV: how it might be extended within medical and nursing education to include more than just courses, and how it might be extended to "applied" ethics teaching more generally.

The third section makes three explicit disclaimers of this dissertation with respect to both evaluating and teaching medical ethics. These disclaimers help define the

scope and limits of the dissertation and, related to this, help eliminate potential misunderstandings.

Buttressing the Model

The "controversies" and "misconceptions" discussed in Chapter II and the tenets of scientism discussed in Chapter III are highly conceptual issues, but are not altogether disconnected from empirical considerations. The three examples--the evaluation findings and the behavior of students and faculty--provide relevant empirical evidence (though by no means direct empirical tests) that may be added to the previous conceptual arguments.

Misconceptions and Controversies Revisited

Controversies

Moral Behavior as a Goal. The ethics in nursing and Focal Problems (1983) examples demonstrated cognitive effects in terms of Knowledge and Reasoning; the Focal Problems (1982) example provided suggestive evidence for the indirect goals. As argued in Chapter II, Knowledge, Reasoning, and the indirect goals make up desired behaviors for the educational context (i.e., behaviors$_e$), and are constitutive of moral behavior in a full-blown sense (i.e., behavior$_r$). The findings from the three examples sanction inferences to the ultimate goal of moral behavior$_r$, and thus buttress the claim of Chapter II that it is unnecessary and

misleading to deny, without qualification, that moral behavior is a goal of medical ethics courses.

Appreciation as a Direct Goal. Appreciation was implicitly adopted in each course as a proximate (i.e., direct) goal. The course designers took student evaluations of their courses very seriously, but measured student evaluations against cognitive goals. The positive attitudes of students accompanied by evidence for cognitive effects in the nursing and Focal Problems (1983) examples were taken to mean that adequate courses had been developed. By contrast, the positive student attitudes not accompanied by good evidence for cognitive effects in the Focal Problems (1982) example prompted changes in the testing and preparation of preceptors. In short, the designers of each of the courses adopted Appreciation as a direct goal for which they were responsible, and construed Appreciation in the way defined in Chapter II, namely, as combining attitudinal and cognitive elements.

Medical Ethics Teaching as Formal. The design and content of the three courses belies the charge that medical ethics teaching is too "formal" (i.e., abstract, irrelevant, and of interest only to philosophers). Each course emphasized problems health care professional inevitably face in the execution of their day-to-day duties, and students (especially the nursing students) and faculty (the Focal Problems preceptors) testified to the relevance and value of the content and approach employed in the three courses.
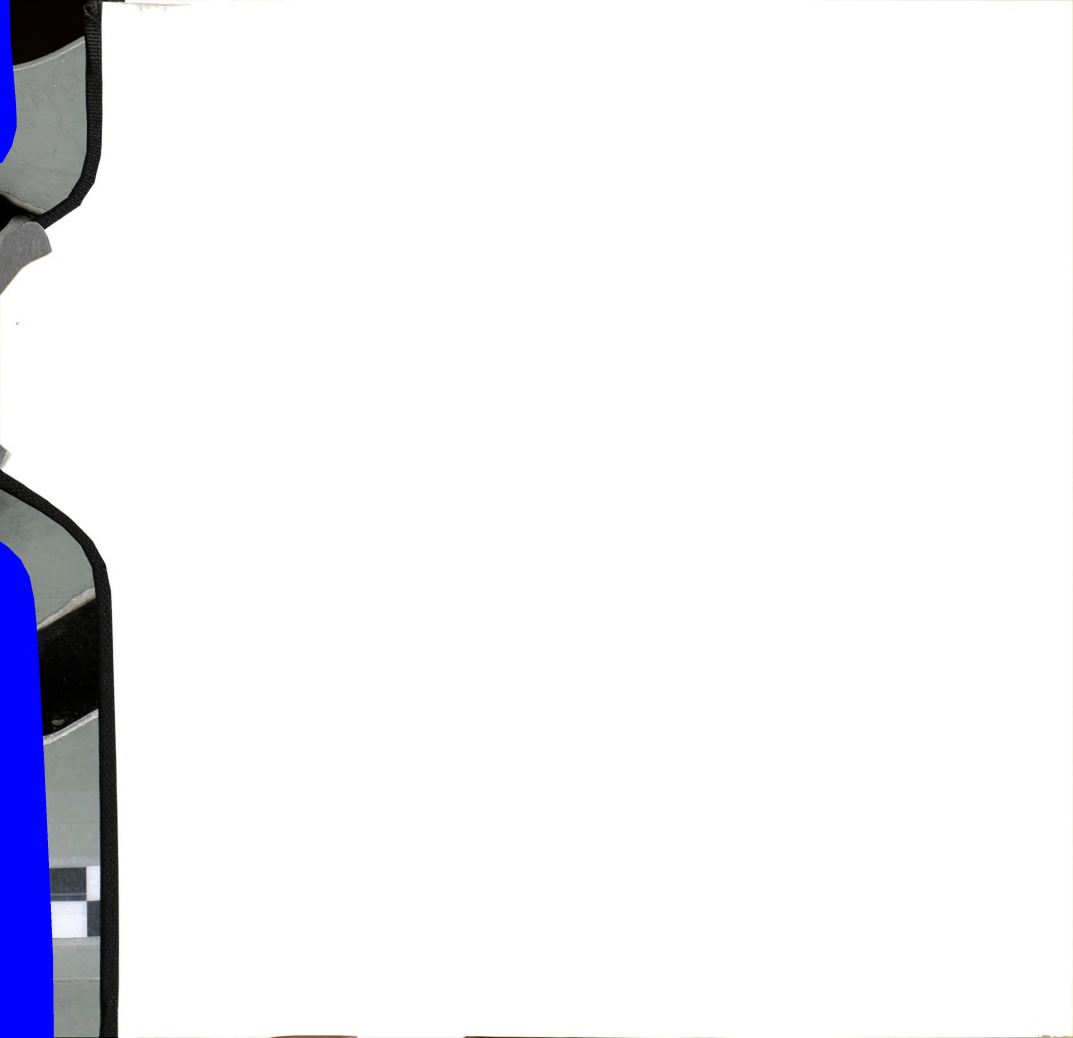
Although medical ethics teaching could be (and perhaps sometimes is) too formal, it does not have to be, simply in virtue of employing a philosophical approach.

<u>Medical Ethics Teaching Incorrectly Downplays Ethical Theory</u>. Because ethical theory is abstract and primarily of interest to philosophers, medical ethics teaching has turned away from stressing ethical theory--teaching ethical theory <u>is</u> teaching that is too "formal". The three course evaluations provide no evidence that greater stress should have been placed on ethical theory. In particular, neither students nor faculty requested such an emphasis. On the contrary, the concentration on concrete cases was praised. Furthermore, there was ample evidence that the courses achieved their desired goals.

<u>Misconceptions</u>

The "misconceptions" discussed in Chapter II were the following: ethics is subsumed by the social sciences; ethics is simply interpersonal skills; professional ethical codes are a sufficient means of dealing with ethical problems; legal considerations preclude ethical ones; ethics teaching conflicts with religious beliefs; ethics teaching is indoctrination; and formal education has no proper role to play in moral education.

If these views in fact apply to medical ethics teaching, one would expect them to be voiced by students and faculty actually exposed to medical ethics teaching. Very little support for these views is to be found on the basis of the
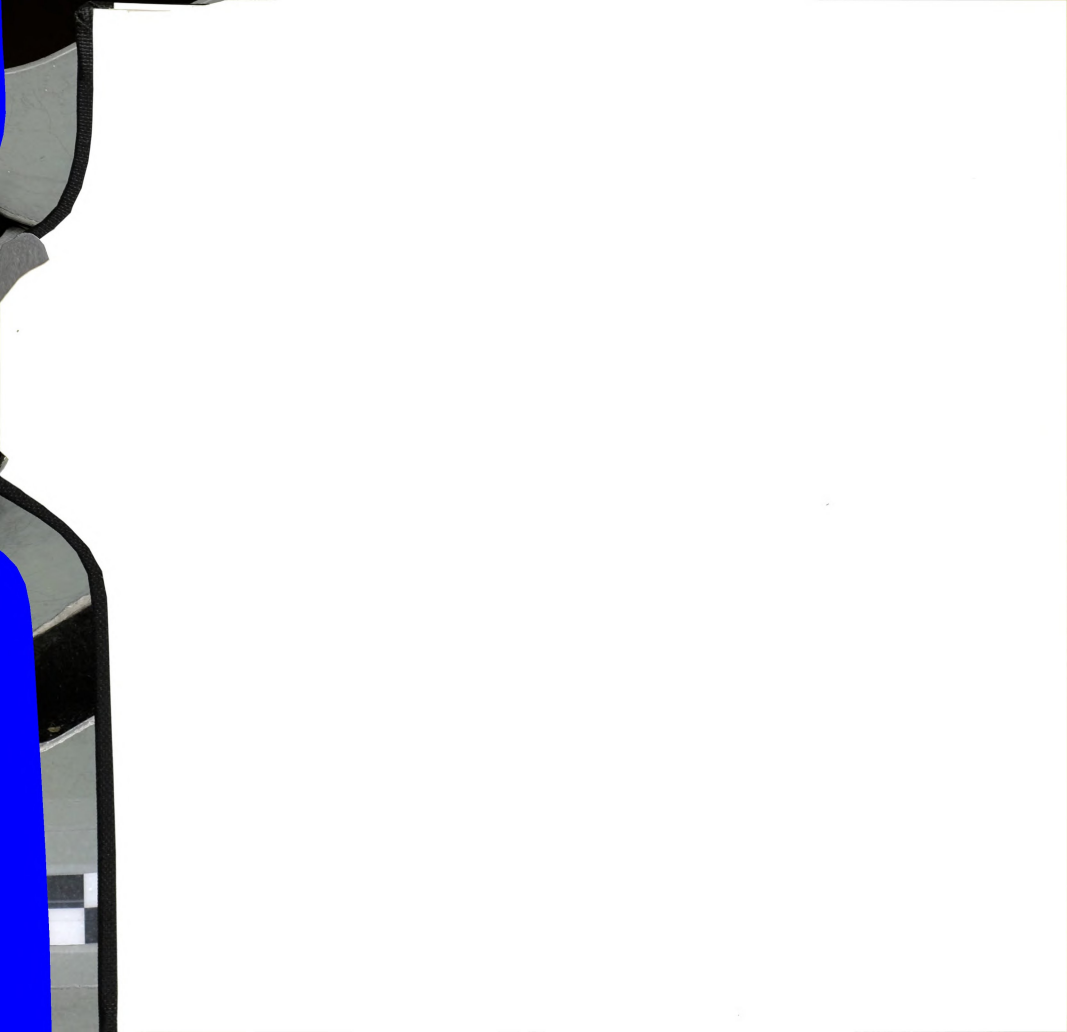
three course evaluations. A few students claimed that the teaching conflicted with religious belief, a few that it was indoctrinating, and a few that testing in ethics is inappropriate because ethics is solely a matter of personal opinion (all from the Focal Problems 1982 example). By and large, however, "misconceptions" (in the form of expressed criticisms of the courses) were uncommon.

The received wisdom is that "misconceptions" about ethics teaching are pervasive. Notwithstanding, students and faculty who had first hand experience with the courses discussed in the preceding chapter did not exhibit "misconceptions". Perhaps "misconceptions" are not as pervasive as they are believed to be, or perhaps the courses in question made major strides in removing them.

Reducing Fallibility Without Scientism

The Fact-Value Distinction

Chapter III argued that defeating the scientistic (positivistic) construal of the fact-value distinction is a pre-condition for making sense of the evaluation of medical ethics teaching—and of educational evaluation in general. The findings from the three examples provide empirical evidence that may be used to further criticize the scientistic fact-value distinction.

If the scientistic construal of the fact-value distinction is correct—if some version of subjectivism is correct in which value judgments have to do only with
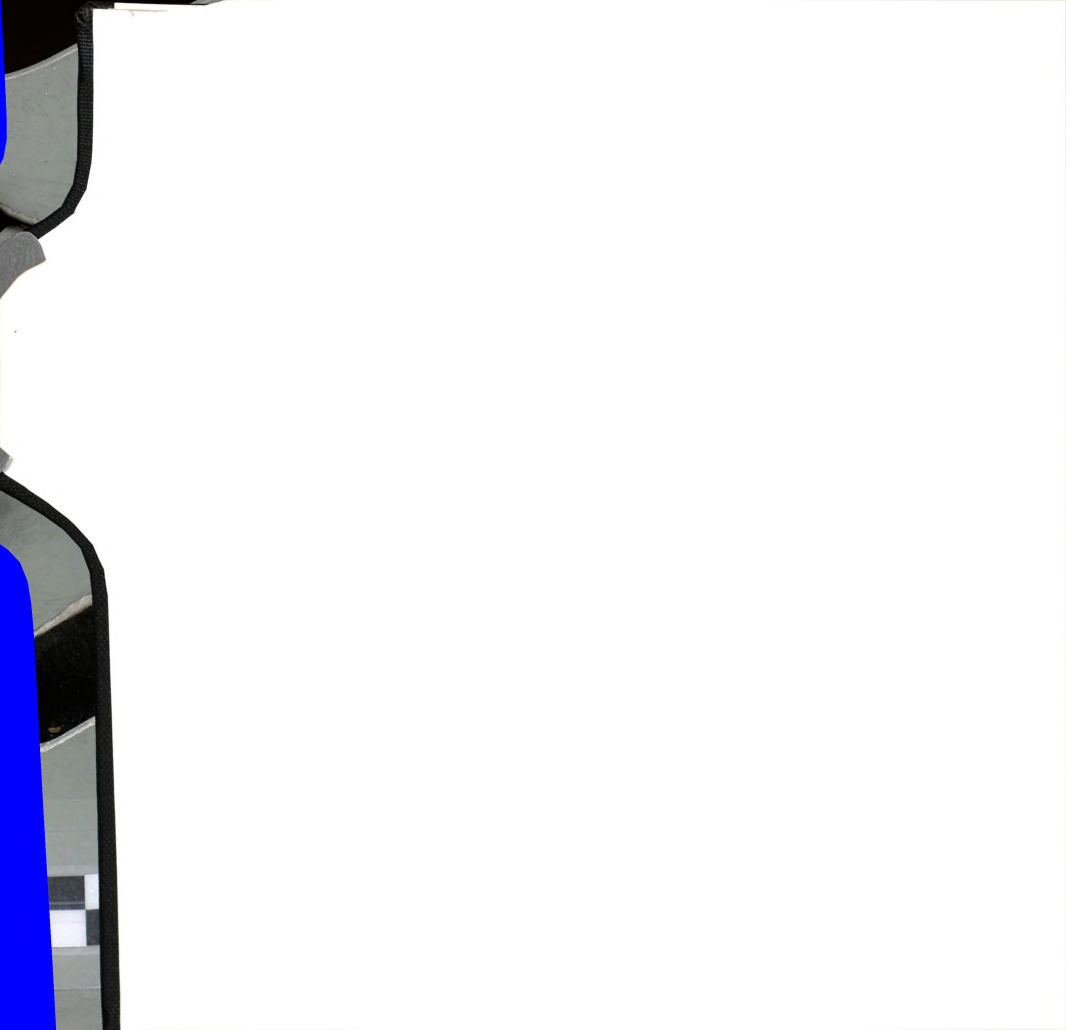
personal, non-cognitive preferences--then two things should
follow: (1) there should be no recognizable cognitive
standards by which to distinguish the warrant of ethical
judgments, and (2) subjectivism should be exemplified in the
behavior of faculty and students. On the other hand, if
these two expectations are contradicted, then the bases for
"value phobia" are further undermined.

1. <u>Cognitive Standards in Ethics</u>. Evidence for
cognitive effects from testing, observation, interviews, or
some combination was present in each of the three examples.
In the nursing and the Focal Problems (1983) examples,
advanced philosophy graduate students rated post-tests
higher than pretests, and did so in terms of accepted
standards of argumentation (e.g., specifying a position,
defending it with reasons, and responding to counter-
arguments). This evidence was buttressed by pre-post
observations in the nursing example, and by observations and
interviews in the Focal Problems (1982) example.

Although in light of these findings one would be hard
pressed to deny that <u>some</u> standards were employed, it might
be objected that these findings only show that students were
brought around to the biases implicit in the content of the
courses. In response to this objection, the
standards--specifying a position, defending it with reasons,
and responding to counter-arguments--speak for themselves as
cognitive standards; they are "biases" only in some very
curious sense. Moreover, the inter-rater agreement
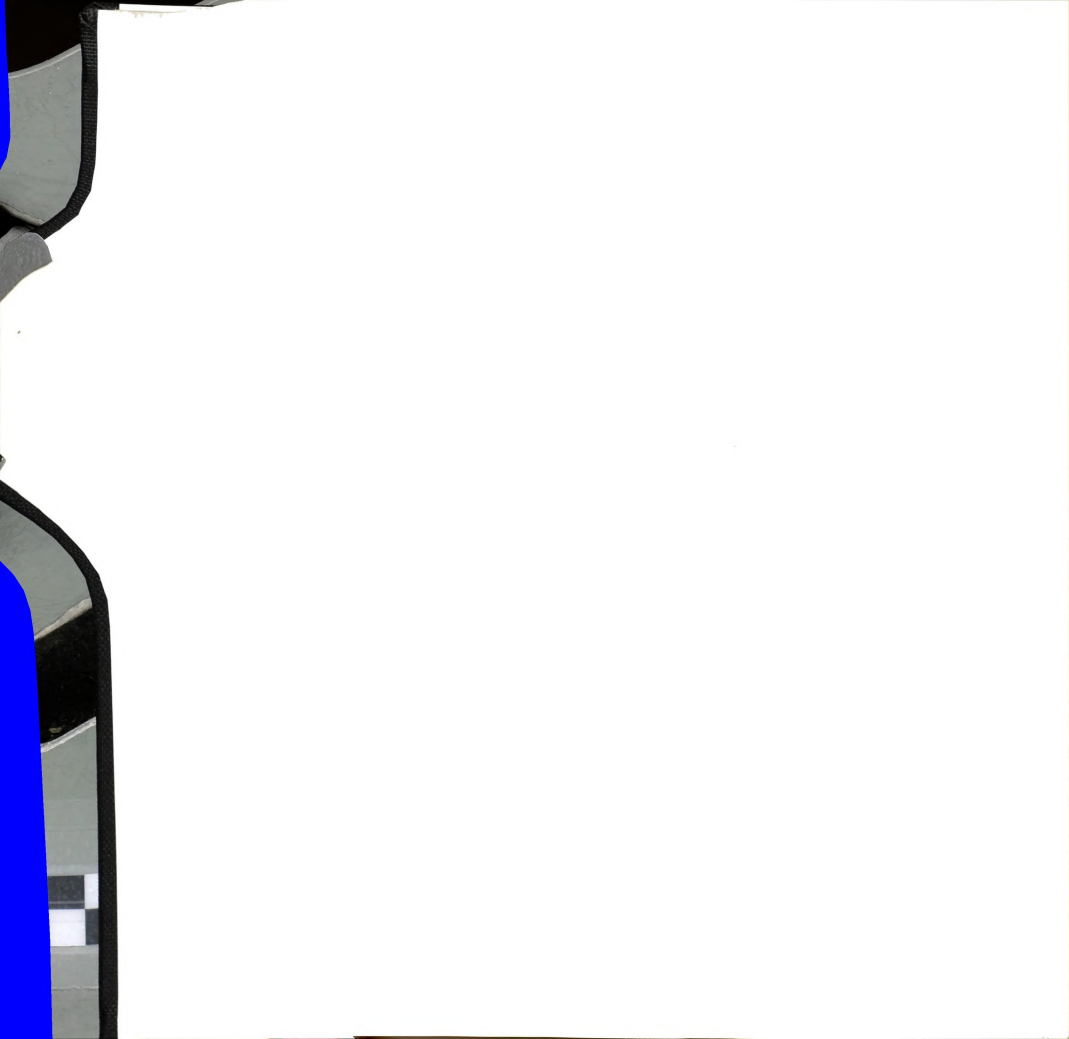
exhibited by the preceptors grading essays in the Focal Problems 1983 example indicates that they were able to recognize and apply the standards in question. It is unlikely that faculty could so easily be converted to accept the "biases" of the course designers.

2. _The Behavior of Faculty and Students_. Students and faculty consistently behaved _as if_ they believed ethical subjectivism is false. That is, data from the three examples regarding discussions and examinations indicates they were engaged in activities that _seemed_ to be substantially cognitive in nature and to involve much more than the mere expressions of emotions and personal preferences. Moreover, as already noted, "misconceptions" in general were rare in the three examples, including the one that formal training in ethics is inappropriate because ethics is essentially non-cognitive. Ethical subjectivism, it seems, is avowed but not lived: the evidence from the three examples supports the contention of Chapter III that ethical subjectivism is not appealing in its own right but follows from the tenets of positivism.

The Quantitative-Qualitative Distinction

The general theoretical points made in Chapter III about the quantitative-qualitative distinction were (1) that the scientistic epistemology which elevates quantitative methods to a distinct and superior position is virtually identical with positivism and is identically flawed, and (2) that

educational evaluation unavoidably makes use of both kinds of data to justify inferences.

The model and undergirding epistemology that guided the three evaluations is supported by the credibility and utilization that the examples enjoyed. In each case qualitative and quantitative methods were combined in a multiple indicators approach which responded to prevailing conditions and purposes, and in each case the findings had sufficient credibility to be utilized. Results of the ethics in nursing and Focal Problems (1983) studies provided evidence that the courses in question were achieving desired and defensible aims, and the results were disseminable to the relevant audiences. Results of the Focal Problems (1982) evaluation raised questions about success and located the source of problems, leading to needed improvements.

At the practical level of the three examples, the scientistic criticisms that the results are suspect or incoherent because qualitative methods were combined with quantitative methods begs the question. The scientistic quantitative-qualitative distinction flows top-down from positivism: in light of the utilization of the results from the three examples, the epistemological standards employed by those interested in information about medical and nursing ethics teaching are not scientistic ones. Saving the scientistic qualitative-quantitative distinction, in addition to meeting the criticisms advanced in Chapter III, thus requires accusing this audience of being ignorant of
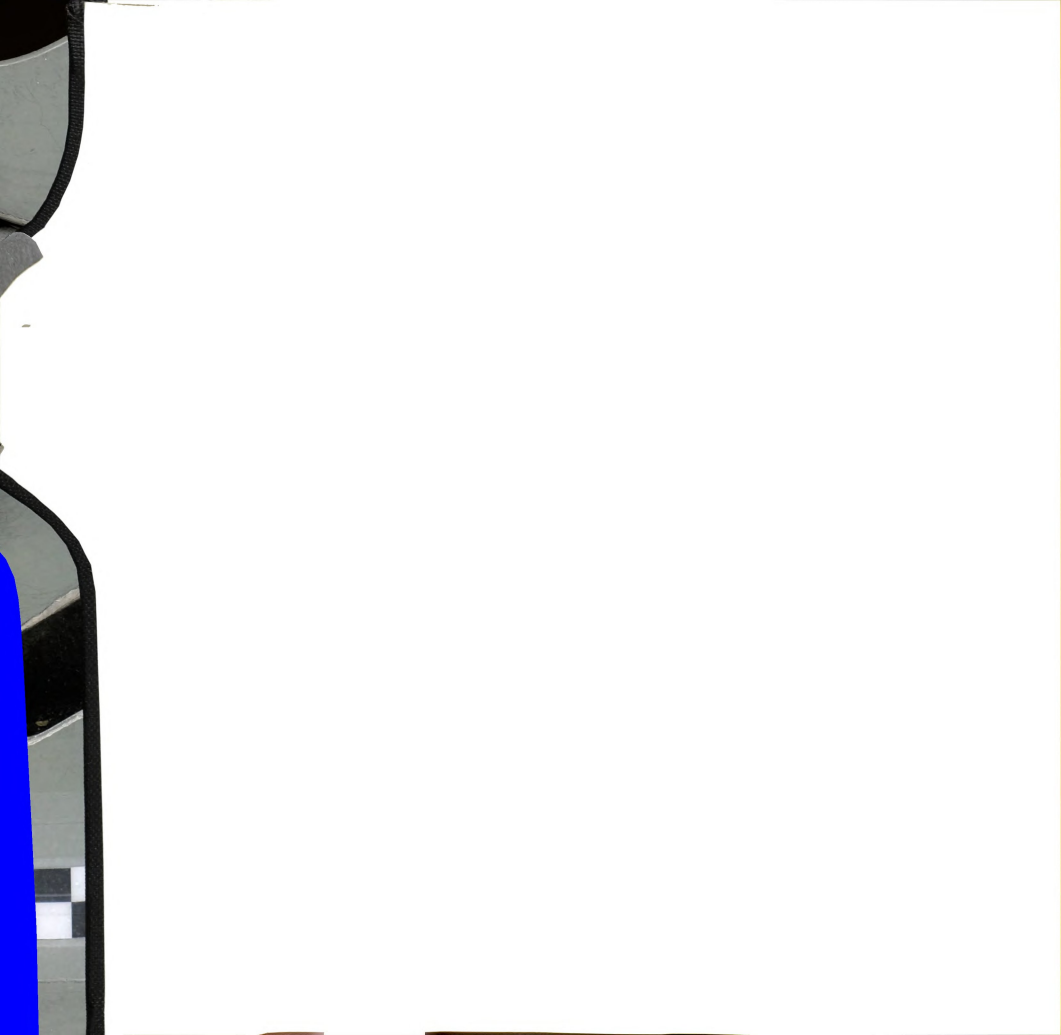
correct epistemology (i.e., it requires accusing them of not knowing a good argument when they see one).

Measurement Grounded in Moral Psychology

Chapter III argued that behavioristic criteria and measures are fundamentally inappropriate for the evaluation of medical ethics teaching and that Kohlbergian criteria and measures are hopelessly flawed from a practical point of view. The relative success of the use of alternative, more intuitive criteria and measures in the three examples reinforces these contentions.

1. Sensitivity to Questions of Interest. The combination of interviews, observations, tests, and questionnaires was sensitive to the goals of Reasoning, Knowledge, and Appreciation. Sensitivity to Reasoning demonstrates the superiority of the strategy of the model employed in the three examples over behaviorism and Kohlberg's theory. Behaviorism fails to even broach the important question of Reasoning (i.e., cognitive moral judgment) and Kohlberg's theory provides an overly formal (if not incorrect and objectionable) construal of it. By contrast, the observational and testing strategies employed in the three examples were able to detect the desired effects. Also, the Focal Problems (1983) example measured Knowledge and Reasoning separately, supporting a distinction in the cognitive aims of ethics teaching that Kohlberg's theory fails to account for.
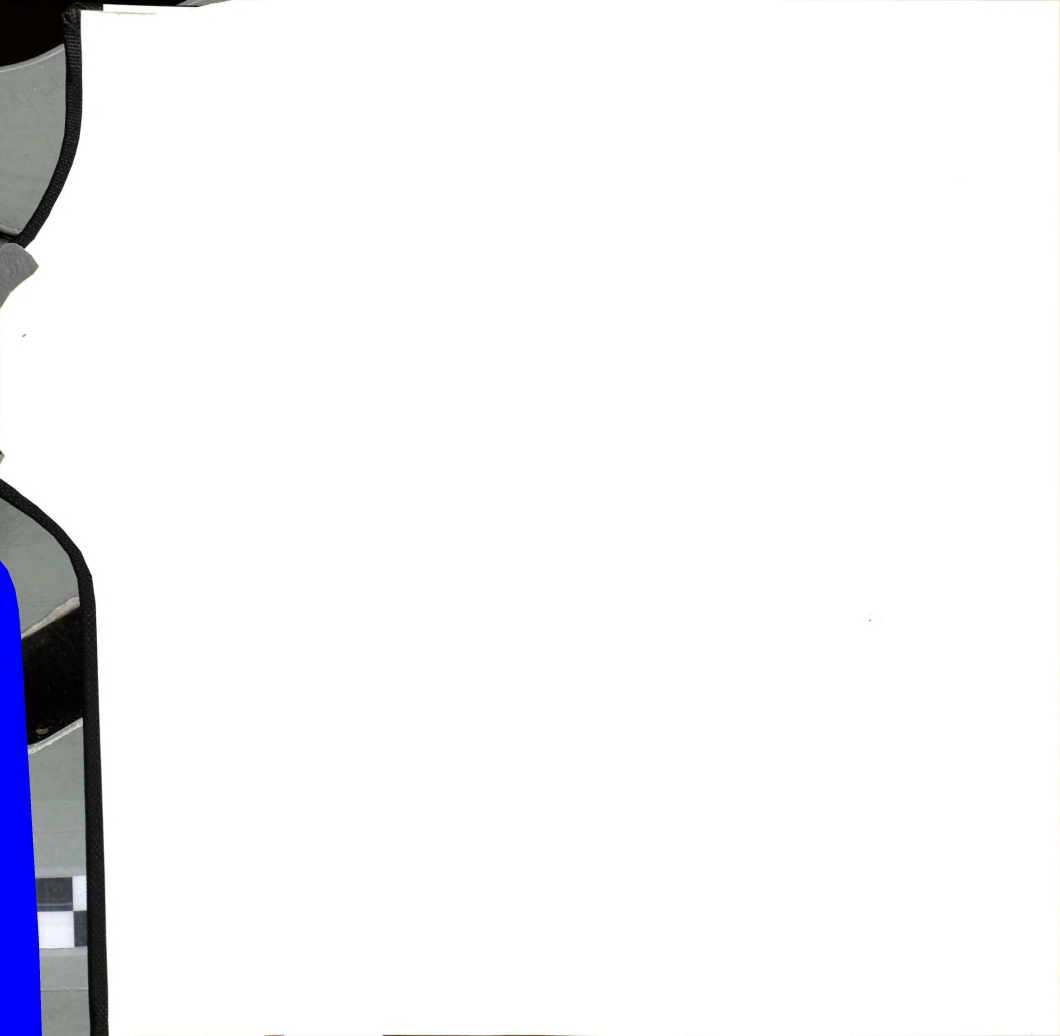
2. <u>Interpretability</u>. The superior sensitivity of the methods of the model employed in the three examples applies perforce to interpretability of results: obtaining no observable effects or observable effects in terms of the wrong criteria is uninformative or misleading. In the nursing and Focal Problems (1983) examples, desired effects were detected in terms of the three major questions of interest, Reasoning, Knowledge, and Appreciation, and the results led the designers to be satisfied with the courses as they stood. In the Focal Problems 1982 example, sources of problems were identified that prompted needed improvements. In none of the examples were there questions in the minds of the designers about what the results of the evaluations meant.

3. <u>Practicability</u>. An approach based on psychologically based measurement requires the efforts of specialists in psychology, who are unlikely to be permanently available to monitor efforts and who employ measures which match their own interests and expertise. Considerable effort is likely to be expended on a one-shot research project that will soon be forgotten (and perhaps never understood or endorsed) by those responsible for teaching. In contrast to this approach, each of the three evaluations discussed in the preceding chapter was accomplished by one staff evaluator working no more than one-quarter time. This time commitment becomes considerably less after strategies and instruments have been developed.

Once deve

can becom

instructo

from expe

medical s

the dual

function

<u>Extendin</u>

For

courses

Interper

indirect

possible

feasible

One

in the c

Universi

format

through

encount

telling

rotatio

disadva

evaluat

Once developed, the kind of measurement strategy employed can become a permanent fixture and turned over to instructors themselves, requiring only periodic monitoring from experts in evaluation (who are usually available in medical schools). The kind of testing employed can serve the dual function of program evaluation and the more common function of evaluating students to assign grades.

Extending the Model

Extending the Model Within Medical Education

For the reasons advanced in Chapter IV, traditional courses are not well-suited for teaching Moral Regard, Interpersonal Skills, Empathy, and Courage, i.e., the indirect goals. However, where direct patient contact is possible, more actively pursuing the indirect goals is both feasible and indicated.

One context for this is the teaching of skills required in the doctor-patient relationship. Southern Illinois University (SIU), for instance, has incorporated a teaching format termed "multiple stations" in which students rotate through several simulated patient encounters. Many of these encounters incorporate common ethical problems like truth telling and informed consent. Another setting is clinical rotations. Hospitals and physicians' offices have the disadvantage of providing less controlled conditions for evaluation than simulated patients. They have the

advantages, however, of a richer variety of encounters and the involvement of actual patients.

It is feasible to incorporate ratings on the indirect goals into evaluations in both contexts. Student performance in the SIU simulated patient encounters is rated in two ways: by the simulated patients themselves and by unobtrusive observers. Incorporating indirect goals into these evaluations, students might be rated on their abilities to "read" patients (Empathy), to follow through in difficult situations (Courage), to demonstrate concern (Moral Regard), and to effectively communicate in a way sensitive to the demands of the situation (Interpersonal Skills). Similar kinds of ratings could be incorporated into actual clinical contexts. Working out the details of evaluation in either context is largely a practical matter. Such evaluations could be accomplished by direct observation resulting in narratives, observational checklists, debriefing interviews, or some combination.

Although feasible, obstacles to this more comprehensive brand of evaluation of medical ethics surely exist. In addition to the practical and moral problems associated with the indirect goals discussed in Chapter II, the simulated patient, physicians' office, and hospital settings introduce imposing practical methodological problems. Moreover, marked resistance exists to what may be perceived as passing judgment on the moral virtue of students. Although, on the one hand, there is a demand to affect moral behavior, on the

other hand, there is a strong presumption against influencing or judging it.

The issue of feasibility, then, cannot be divorced from the general milieu of medical education. Given the recalcitrant nature of the traits and dispositions that correspond to the indirect goals, medical schools need to incorporate these traits and dispositions as explicit criteria in admissions standards. This would both alleviate some of the potential pedagogical problems and clarify the commitment to moral behavior as essential to the practice of medicine. An explicit statement and commitment is also required within curricula. Only then will it be reasonable to demand that medical ethics teaching promote the desired aims and be accountable for them. And only then will evaluation of such aims be appropriate and not be met with considerable resistance from all sides.

Extending the Model to Applied Ethics Generally

The Wilson+1 goals are specific to medical ethics only with regard to particular ethical issues of interest and the relationship between such ethical issues and appropriate domains of knowledge. With suitable adjustments in the specific issues addressed, the methods of evaluating medical ethics teaching advanced in this dissertation may be straightforwardly implemented in other areas of applied ethics.

There are roughly two major types of educational programs in which applied ethics plays (or should play) a

role. One kind is the standard undergraduate program, consisting of a variety of courses leading to a undergraduate degree. Engineering, business, and journalism are examples of such programs. Applied ethics teaching in these areas would seem limited to self-standing courses, and the model for evaluating medical ethics courses of Chapter IV is readily adaptable to the differing content and issues involved. A second kind of program is one that adds apprenticeship field experience to standard course work. In addition to medicine and nursing, veterinary medicine, dentistry, and primary and secondary teaching are examples of this kind of program. The nature of these educational programs, as well as the professions themselves, entail direct interpersonal contact with individual persons. This kind of training program is amenable to the broad model which includes both direct and indirect goals and is subject to the same limitations and problems mentioned in connection with medical education.

Disclaimers

This dissertation does not have the presumptuous aim of providing the last word on how medical ethics ought to be evaluated, much less on how it ought to be taught. In order to avoid possible misunderstandings on these two points, this section makes three explicit disclaimers.

Disclaimer 1. To borrow from Clouser (1980), evaluation should not be the tail that wags the dog. If certain
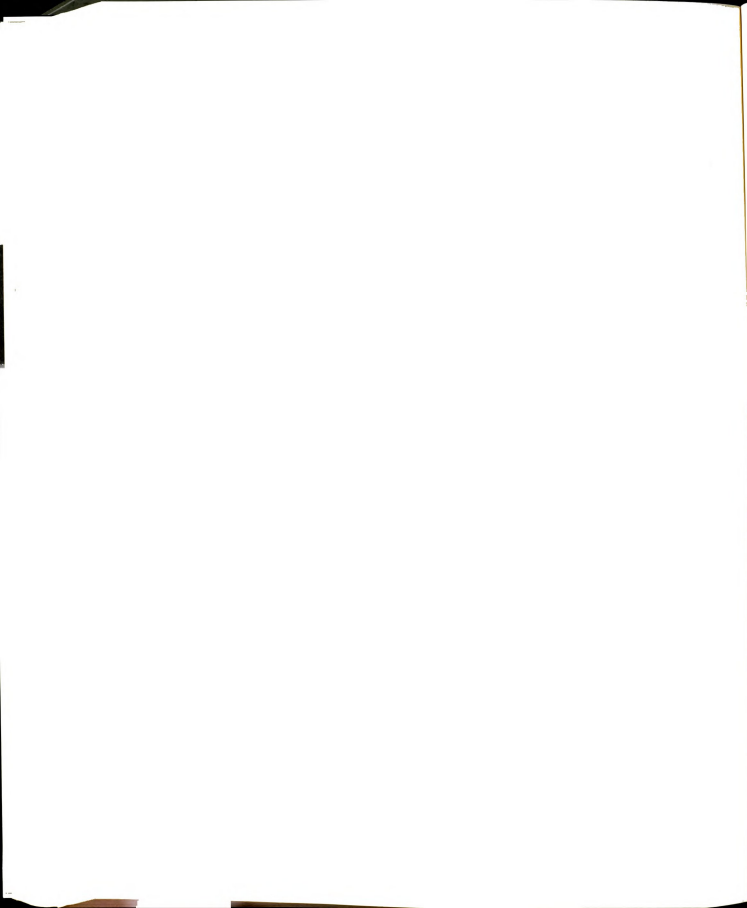
pedagogical practices that impinge on evaluation are defensible on independent grounds (for instance, requiring students to write papers), such practices should not be avoided in deference to the aims and preferred methods of evaluation research. To be sure, evaluators have something to say about legitimate aims and means of achieving them, but aims and means should not be compromised merely to simplify evaluation or to yield more methodologically defensible evaluation results. As stated previously, evaluation should work its way out from the nature of the object under investigation--this is a fundamental tenet of functional evaluation.

Disclaimer 2. This dissertation aimed to show that medical ethics teaching can be evaluated in a defensible way. This is a much more modest aim than showing the only or best way to do evaluation.

The basic functional model used in Chapter IV was dictated by resources and practices that prevailed at a given institution at a given point in time, and one could easily imagine both being different. For example, literature might be incorporated into a medical ethics course to, among other things, improve (versus elicit) students ability to empathize. Incorporating literature in this way would press the classification of "eliciting Empathy" as an indirect (or process) goal. If resources were available to carefully research the impact of literature (e.g., students might be given passages from a

novel pre-post and asked on each occasion to describe the feelings of characters), the incorporation of literature would also render the basic functional model of Chapter IV methodologically inadequate. To take another example, one defensible aim of medical ethics teaching is helping future physicians balance their own interests against the interests of others (e.g., Pellegrino et al., 1985). Given this aim of medical ethics teaching, it could be suggested that the Wilson+1 goals are incomplete.

The criticisms of the preceding paragraph would undermine this dissertation if the aim was to provide a blueprint for doing evaluation, but the aim was merely to show that methodologically defensible, yet fruitful, evaluation is possible. For the reasons given in the introductory chapter, this is something that needed showing.

Disclaimer 3. Just as this dissertation aimed to show only that medical ethics teaching can be evaluated, it aimed, in the process, to show only that the three courses used as examples achieved desirable goals. Showing that these courses achieved desirable goals falls considerably short of showing any of them is the best way to teach medical ethics. Determining what methods of teaching medical ethics are best would require careful comparative studies that go far beyond the limitations of this dissertation.

Closing Remarks

Ethics is frequently viewed by educational evaluators as different from other academic subjects. The perceived difference finds its roots in the <u>great divide</u> between what is purportedly objective, scientific, factual, quantitative, and cognitive on the one hand and what is subjective, unscientific, value-laden, qualitative, and affective on the other. Ethics is believed by many to fall squarely within the second set of descriptions, which gives rise to "value phobia" and the notion that evaluating ethics in terms of cognitive standards is out-of-place. Even those who stop short of going all the way with this demand that cognitive standards have some objective basis in science, most notably, in theories of cognitive moral development.

For their part, ethics instructors also have a tendency to view ethics as radically different from other academic subjects, but for different reasons than educational evaluators. They believe ethics is not amenable to reductivist social research methods and is something that must be valued for its own sake. They identify all evaluation as reductivist and draw the conclusion that their teaching is not subject to the kind of methods that experts in evaluation employ.

This dissertation has shown that evaluators and medical ethics instructors who hold the above views are mistaken. Educational evaluation incorporates values by its very nature, and ethics differs from other objects of evaluation

only as a matter of degree. Ethics teaching thus does not defy evaluation because it inherently subjective, unscientific, or anything of that sort. The other side of this coin is that ethics teaching does not lie above the fray when it comes to demonstrations of worth and effectiveness.

REFERENCES

Aleamoni, L. (1981). Student ratings of instruction. In J. Millman (Ed.), *Handbook of Teacher Evaluation* (pp. 110-145). Beverly Hills, California: Sage Publications.

American Association of American Medical Colleges (1984). *Physicians for the Twenty-First Century*. Washington, D.C.

Anderson, S. and Ball, S. (1978). *The Profession and Practice of Program Evaluation*. San Francisco: Jossey-Bass.

Archambault, A. (1975). Criteria for success in moral education. In I. Chazan and J. Soltis (Eds.), *Moral Education* (pp. 159-69). New York: Teachers College Press.

Beauchamp, T. (1982). What philosophers can offer. *Hastings Center Report*, *12*(3), 13-14.

Beauchamp, T. and Walters, L. (1978). *Contemporary Issues in Bioethics*. Belmont, California: Wadsworth Publishing Co.

Benjamin, M. and Curtis, J. (1981). *Ethics in Nursing*. New York: Oxford University Press.

Bok, S. (1978). *Lying: Moral Choice in Public Policy and Private Life*. New York: Pantheon Books.

Callahan, D. (1980). Goals in teaching ethics. In D. Callahan and S. Bok (Eds.), *Ethics Teaching in Higher Education* (pp. 61-74). New York: The Hastings Center.

Callahan, D. (1978). The rebirth of ethics. *National Forum*, *58*(2), 9-12.

Campbell, D. (1982). Experiments as arguments. In E. House (Ed.), *Evaluation Studies Review Annual* (pp. 117-28). Beverly Hills, California: Sage Publications.

Campbell, D. (1979). "Degrees of freedom" and the case study. In T. Cook and C. Reichardt (Eds.), *Qualitative and Quantitative Methods in Evaluation Research* (pp. 49-67). Beverly Hills, California: Sage Publications.

Campbell, D. (1974). Qualitative knowing in action research. Kurt Lewin Award Address, Society for the Psychological Study of Social Issues, meeting with the American Psychological Association, New Orleans, September.

Caplan, A. (1983). Ethical engineers need not apply: the state of applied ethics today. In S. Gorovitz et al. (Eds.), Moral Problems in Medicine (2nd Ed.), (pp. 38-43). Englewood Cliffs, New Jersey: Prentice-Hall.

Caplan, A. (1980). Evaluation and the teaching of ethics. In D. Callahan and S. Bok (Eds.), Ethics Teaching in Higher Education (pp. 133-50). New York: The Hastings Center.

Clements, C. and Sider R. (1983). Medical ethics assault on medical values. Journal of the American Medical Association, 250(15), 2011-15.

Clouser, K. D. (1980). Teaching Bioethics: Strategies, Problems amd Resources. New York: The Institute of Society, Ethics and Life Sciences.

Clouser, K. D. (1973). Medical ethics: some realistic expectations. Journal of Medical Education, 48, 373-74.

Cronbach, L. (1982). Designing Evaluations of Educational and Social Programs. San Francisco: Jossey-Bass.

Cronbach, L. and Associates (1980). Toward Reform of Program Evaluation. San Francisco: Jossey-Bass.

Culver, C. et al. (1985). Basic curricular goals in medical ethics. New England Journal of Medicine, 312, 253-56.

Daniels, N. (1979). Wide reflective equilibrium and theory acceptance in ethics. Journal of Philosophy, May, 256-82.

Dewey, J. (1944). Democracy and Education. New York: The Free Press.

Dewey, J. (1939). Theory of Valuation. Chicago: University of Chicago Press.

Dykstra, K. (1981). Vision and Character. New York: Paulist Press.

Elstein, A., Sprafka, S. and Shulman, L. (1978). Medical Problem Solving: An Analysis of Clinical Reasoning. Cambridge, Massachussets: Harvard University Press.

Flanagan, O. (1982). Some philosophical reflections on the moral psychology debate. Ethics, 92(3), 499-512.

Frankena, W. (1975). Toward a philosophy of moral education. In I. Chazan and J. Soltis (Eds.), Moral Education (pp. 148-58). New York: Teachers College Press.

Gilligan, C. (1977). In a different voice: women's conception of the self and of morality. Harvard Educational Review, 47, 481-517.

Good, I. J. (1983). The Bayesian influence, or how to sweep subjectvism under the carpet. In Good Thinking (pp. 22-58). Minneapolis: The University of Minnesota Press.

Goodpastor, K. (1982) Is teaching ethics 'making' or 'doing'?, Hastings Center Report 12(1), 37-39.

Gorovitz, S. (1984). Preparing for the perils of practice. Hastings Center Report, 14(6), 38-41.

Hall, R. and Davis J. (1975). Moral Education in Theory and Practice. Buffalo, New York: Prometheus Books.

Hart, H. L. A. (1961). The Concept of Law. London: Oxford University Press.

Hastings Center Staff (1980). The teaching of ethics in American higher education: an empirical synopsis. In D. Callahan and S. Bok (Eds.), Ethics Teaching in Higher Education (pp. 153-70). New York: Plenum Press.

Howe, K. and Jones, M. (1984). Techniques for evaluating student performance in a preclinical medical ethics course. Journal of Medical Education, 59, 350-52.

Howe, K., Holmes, M. and Elstein, A. (1984). Teaching clinical decision making. Journal of Medicine and Philosophy, 9(2), 215-228.

Howe, K. (1982). Evaluating philosophy teaching: assessing student mastery of philosophical objectives in nursing ethics. Teaching Philosophy, 5(1), 11-22.

Hunt, R. and Aras, J. (Eds.), (1977). Ethical Issues in Modern Medicine. Palo Alto, California: Mayfield Publishing Company.

Iozzi, L. and Paradise-Maul, J. (1980). Issues at the interface of science, technology, and society. In L. Kuhmerker et al. (Eds.), Evaluating Moral Development and Evaluating Educational Programs That Have a Moral Dimension. Schenectady, New York: Character Research Press.

Jensen, A. (1984). Political ideologies and educational research. Phi Delta Kappan, 65(7), 460-63.

Jones, M. and Howe, K. (1983). Measuring analytic moral reasoning in medical students. Unpublished.

Kahn, G., Cohen, B. and Jason, H. (1979). The teaching of interpersonal skills in U. S. medical schools. *Journal of Medical Education*, 54(1), 29-35.

Kohlberg, L. (1982). A reply to Owen Flanagan and some comments on the Puka-Goodpastor exchange. *Ethics*, 92(3), 513-28.

Krebs, D. (1982). Psychological approaches to altruism: an evaluation. *Ethics*, 92(3), 447-58.

Kuhn, T. (1961). The function of measurement in modern physical science. In H. Wolf (Ed.), *Quantification*, (pp. 31-63). New York: Bobbs-Merril Company.

Lickona, T. (1980). What does moral psychology have to say to the teacher of ethics? In S. Bok and D. Callahan (Eds.), *Ethics Teaching in Higher Education* (pp. 103-32). New York: Plenum Press.

Lipman, M., Sharp, A. and Oscanyan, F. (1977). *Philosophy in the Classroom*. West Caldwell, New Jersey: Universal Diversified Services, Inc.

MacIntyre, A. (1981). *After Virtue*. Notre Dame, Indiana: Notre Dame University Press.

MacIntyre, A. (1980). A crisis in moral philosophy: why is the search for the foundations of ethics so frustrating? In H. T. Engelhardt and D. Callahan (Eds.), *Knowing and Valuing* (pp.18-35). New York: Institute of Society, Ethics and Life Sciences.

Mackenzie, B. (1977). *Behaviorism and the Limits of Scientific Method*. Atlantic Highlands, New Jersey: Humanities Press Inc.

Macklin, R. (1980). Problems in the teaching of ethics: pluralism and indoctrination. In S. Bok and D. Callahan (Eds.), *Ethics Teaching in Higher Education* (pp. 81-102). New York: Plenum Press.

McPeck, J. (1980). *Critical Thinking in Education*. New York: St. Martin's Press.

Melden, A. I. (1966). Free actions. In B. Berofsky (Ed.) *Free Will and Determinism* (pp. 198-220). New York: Harper and Row Publishing Company.

Mercer, J. (1971). Institutionalized anglocentrism: labeling mental retardates in the public schools. In P. Orlenas and W. Ellis (Eds.), Race, Change, and Urban Society (pp. 311-38). Beverly Hills, California: Sage Publications.

Messick, S. (1981). Evidence and ethics in the evaluation of tests. Educational Researcher, 10(9), 9-20.

Nobel, C. (1982). Ethics and experts. Hastings Center Report, 12(3), 7-9.

Noonan, J. (1977). An almost absolute value in history. In R. Hunt and J. Aras (Eds.), Ethical Issues in Modern Medicine. Palo Alto, California: The Mayfield Publishing Company.

Nowell-Smith, P. H. (1967). Religion and morality. In Paul Edwards (Ed.), The Encyclopedia of Philosophy, Vol. 7, (pp. 150-58). New York: Macmillan Free Press.

Patton, M. (1983). Designing Evaluations of Educational and Social Programs by L. Cronbach. Evaluation News, 4(2), 29-33.

Patton, M. (1980). Qualitative Evaluation Methods. Beverly Hills, California: Sage Publications Inc.

Pellegrino, E., Hart, R., Henderson, S., Loeb, S., and Edwards, G. (1985). Relevance and utility of courses in medical ethics. Journal of the American Medical Association, 253, 49-53.

Peters, R. S. (1967). Ethics and Education. Atlanta, Georgia: Scott, Foresman and Company.

Phillips, D.C. (1983). After the wake: postpositivistic educational thought. Educational Researcher, 12(5), 4-12.

President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research (1982). Making Health Care Decisions. Washington, D.C.: U.S. Government Printing Office.

Puka, B. (1982). An interdisciplinary treatment of Kohlberg. Ethics, 92(3), 468-90.

Putnam, H. (1983). How not to solve ethical problems. Lindley Lectures. Lawrence, Kansas: University of Kansas.

Quine, W.V.O. (1970). The basis of conceptual schemes. In The Foundations of Knowledge C. Landesman (Ed.), (pp. 160-72). Englewood Cliffs, New Jersey: Prentice Hall, Inc.

Quine, W.V.O. (1969a). Epistemology naturalized. In
  Ontological Relativity and Other Essays, (pp. 69-90). New
  York: Columbia University Press.

Quine, W.V.O. (1969b). Natural kinds. In _Ontological
  Relativity and Other Essays_, (pp. 114-38). New York:
  Columbia University Press.

Quine, W.V.O. (1962). Two dogmas of empiricism. In _From a
  Logical Point of View_ (2nd. ed.). Cambridge, Mass.:
  Harvard University Press.

Rachels, J. (1975). Active and passive euthanasia. _New
  England Journal of Medicine_, 292, 78-80.

Rawls, J. (1971). _A Theory of Justice_. Cambridge, Mass.: The
  Belknap Press of Harvard University Press.

Reichardt, C. and Cook, T. (1979). Beyond qualitative
  versus quantitative methods. In T. Cook and C. Reichardt
  (Eds.), _Qualitative and Quantitative Methods in Evaluation
  Research_ (pp. 7-32). Beverly Hills, California: Sage
  Publications.

Rest, J. (1982). A psychologist looks at the teaching of
  ethics. _Hastings Center Report_, _12_(1), 29-36.

Rest, J. (1979). _Development in Judging Moral Issues_.
  Minneapolis, Minnesota: University of Minnesota Press.

Rorty, R. (1982a). Method, social science and social hope.
  In _Consequences of Pragmatism_ (pp. 191-210). Minneapolis:
  University of Minnesota Press.

Rorty, R. (1982b). Pragmatism relativism and irrationalism.
  In _Consequences of Pragmatism_ (pp. 160-75). Minneapolis:
  University of Minnesota Press.

Rorty, R. (1979). _Philosophy and the Mirror of Nature_.
  Princeton, New Jersey: Princeton University Press.

Ruddick, W. (1981). Can doctors and philosophers work
  together? _The Hastings Center Report_, _11_(2), 12-17.

Rushton, P. (1982a). Altruism and society: a social learning
  perspective. _Ethics_, _92_(3), 425-46.

Rushton, P. (1982b). Moral cognition, behaviorism and social
  learning theory. _Ethics_, _92_(3), 459-67.

Ryle, G. (1949). _The Concept of Mind_. London: Hutchinson and
  Company, Ltd.

Scheffler, I. (1978). _The Language of Education_.
Springfield, Illinois: Charles C. Thomas.

Scriven, M. (1983). The evaluation taboo. In E. House (Ed.)
Philosophy of Evaluation (pp. 75-82). San Francisco:
Jossey-Bass.

Scriven, M. (1981). Summative teacher evaluation. In J.
Millman (Ed.), _Handbook of Teacher Evaluation_ (pp.
244-71). Beverly Hills, California: Sage Publications.

Scriven, M. (1979). Clinical judgment. In H. T. Engelhardt,
S. Spicker, and B. Towers (Eds.), _Clinical Judgment: A
Critical Appraisal_ (pp. 3-16). Boston: Reidel Publishing
Company.

Scriven, M. (1975). Cognitive moral education. _Phi Delta
Kappan_, _56_, 680-94.

Scriven, M. (1973). The methodology of evaluation. In B.
Worthen and J. Sanders (Eds.), _Educational Evaluation:
Theory and Practice_ (pp. 60-103). Belmont, California:
Wadsworth Publishing Company.

Scriven, M. (1972b). Objectivity and subjectivity in
educational research. In L. Thomas and H. Richey (Eds.),
Philosophical Redirection of Educational Research, The
Seventy-first Yearbook of the National Society for the
Study of Education (pp. 94-142). Chicago: University of
Chicago Press.

Scriven, M. (1969). Logical positivism and the behavioral
sciences. In P. Achenstein and S. Barker (Eds.), _The
Legacy of Logical Positivism_ (pp. 195-210). Baltimore: The
John Hopkins Press.

Shaw, A. (1977). Dilemmas of "informed consent" in children.
In R. Hunt and J. Aras (Eds.), _Ethical Issues in Modern
Medicine_, (pp. 182-95). Palo Alto, California: Mayfield
Publishing Companny.

Sider, R. and Clements, C. (1984). Medical ethics assault on
medical values. _Journal of the American Medical
Association_, _251_(21), 2791-94.

Siegler, M., Rezler, A., and Connell, K. (1982). Using
simulated case studies to evaluate a clinical ethics
course for junior students. _Journal of Medical Education_,
_57_, 380-85.

Singer, M. (1970). Generalization in ethics. In W. Sellars
and J. Hospers (Eds.), _Readings in Ethical Theory_,
(529-47). New York: Appleton-Century-Crofts, Educational
Division, Meredith Corporation.

Singer, P. (1982). How do we decide? <u>Hastings Center Report</u>, <u>12</u>(3), 9-11.

Smith, J. K. (1983a). Quantitative versus qualitative research: an attempt to clarify the issue. <u>Educational Researcher</u>, <u>12</u>(3), 6-13.

Smith, J. K. (1983b). Quantitative versus interpretive: the problem of conducting social inquiry. In E. House (Ed.), <u>Philosophy of Evaluation</u> (pp. 5-27). San Francisco: Jossey-Bass.

Stolman, C. and Doran, R. (1982). Development and validation of a test instrument for assessing value preferences in medical ethics. <u>Journal of Medical Education</u>, <u>57</u>, 170-79.

Starr, P. (1982). <u>The Social Transformation of American Medicine</u>. New York: Basic Books.

Taylor, C. (1964). <u>The Explanation of Behaviour</u>. New York: The Humanities Press.

Toulmin, S. (1981). The tyranny of principles. <u>The Hastings Center Report</u>, <u>11</u>(6), 31-39.

Toulmin, S. (1960). <u>The Philosophy of Science</u>. New York: Harper and Row Publishers.

Troyer, J. (1982). Ethics in nursing, by Martin Benjamin and Joy Curtis. <u>Journal of Medicine and Philosophy</u>, <u>7</u>(4), 382-84.

Veatch, R. (1977). <u>Case Studies in Medical Ethics</u>. Cambridge, Massachussets: Harvard University Press.

Wikler, D. (1982). Ethicists, critics and expertise. <u>The Hastings Center Report</u>, <u>12</u>(3), 12-13.

Wilson, J. (1983). A letter from Oxford. <u>Harvard Educational Review</u>, <u>53</u>(2), 190-94.

Wilson, J. (1973). <u>A Teacher's Guide to Moral Education</u>. London: Geoffrey Chapman.

Wilson, J. (1969). <u>Moral Education and the Curriculum</u>. Oxford: Pergammon Press.

Wilson, J. (1967). What is moral education? Part 1. In J. Wilson, N. Williams, and B. Sugarman, <u>Introduction to Moral Education</u> (pp. 39-226). Baltimore: Penguin Books.

Wittgenstein, L. (1958). <u>The Philosophical Investigations</u>. New York: The Macmillan Company.

Worthen, B. and Sanders, J. (1973). _Educational Evaluation: Theory and Practice_. Belmont, California: Wadsworth Publishing Company.

Wren, T. (1982). Social learning theory, self-regulation, and morality. _Ethics, 92_(3), 409-24.

APPENDICES

APPENDIX A

ETHICS IN NURSING INSTRUMENTS

Nursing Student Questionnaire

Key: SA: Strongly agree
A: Agree
N: Neither agree nor disagree
SD: Strongly disagree

		SA	A	N	D	SD
1.	The course improved my ability to see the complexity of moral problems which face nurses.	__	__	__	__	__
2.	The course helped me develop a framework or basis for my moral positions.	__	__	__	__	__
3.	The course helped me see the importance of giving reasons and careful arguments for my moral beliefs and decisions.	__	__	__	__	__
4.	The topics discussed and the written assignments were relevant to nursing.	__	__	__	__	__
5.	The cases discussed were realistic enough to bring out the emotional side of moral problems.	__	__	__	__	__
6.	Discussing cases is a better way to learn about moral problems in nursing than lectures.	__	__	__	__	__
7.	I will be a better nurse because I have had this course.	__	__	__	__	__
8.	My patients will receive better care because I have had this course.	__	__	__	__	__
9.	I feel more confident about recognizing and dealing with moral problems because I have had the course.	__	__	__	__	__
10.	All medical professionals should study moral problems in medicine in a similar way.	__	__	__	__	__
11.	Multiple choice or true-false exams would have been a better way of grading than essays.	__	__	__	__	__
12.	Writing the essays contributed to what I learned in the course.	__	__	__	__	__

A1

13. The instructors' comments on my written   ___ ___ ___ ___ ___
    assignments helped me learn.

14. The course improved my <u>general</u> abilities to   ___ ___ ___ ___ ___
    discuss, reason and write about issues, not
    just my abilities with respect to moral
    problems in nursing.

APPENDIX B

FOCAL PROBLEMS (1982) INSTRUMENTS

1. Preceptors' Names_____

2. How important is it for physicians to be skilled in dealing with
    ethical problems?

    _____quite important

    _____important

    _____not very important

3. Will your experience in this course help you deal with such
    problems in your practice?

    _____yes

    _____no

    _____don't know

    Comments:_____

    _____

4. Were the aims of the course initially unclear? ___Yes ___No Did

    they become clearer as the term progressed? ___Yes ___No Did you

    get a "feel" for the issues and how to deal with them? ___Yes ___No

    If not, what would have helped you?_____

    _____

5. Please rate the following course areas.

	low 1	2	average 3	4	high 5
usefulness of lectures	___	___	___	___	___
interest in lectures	___	___	___	___	___
usefulness of discussion groups	___	___	___	___	___
interest in discussion groups	___	___	___	___	___
quality of readings	___	___	___	___	___
quality of discussion guides	___	___	___	___	___
skill of preceptors	___	___	___	___	___
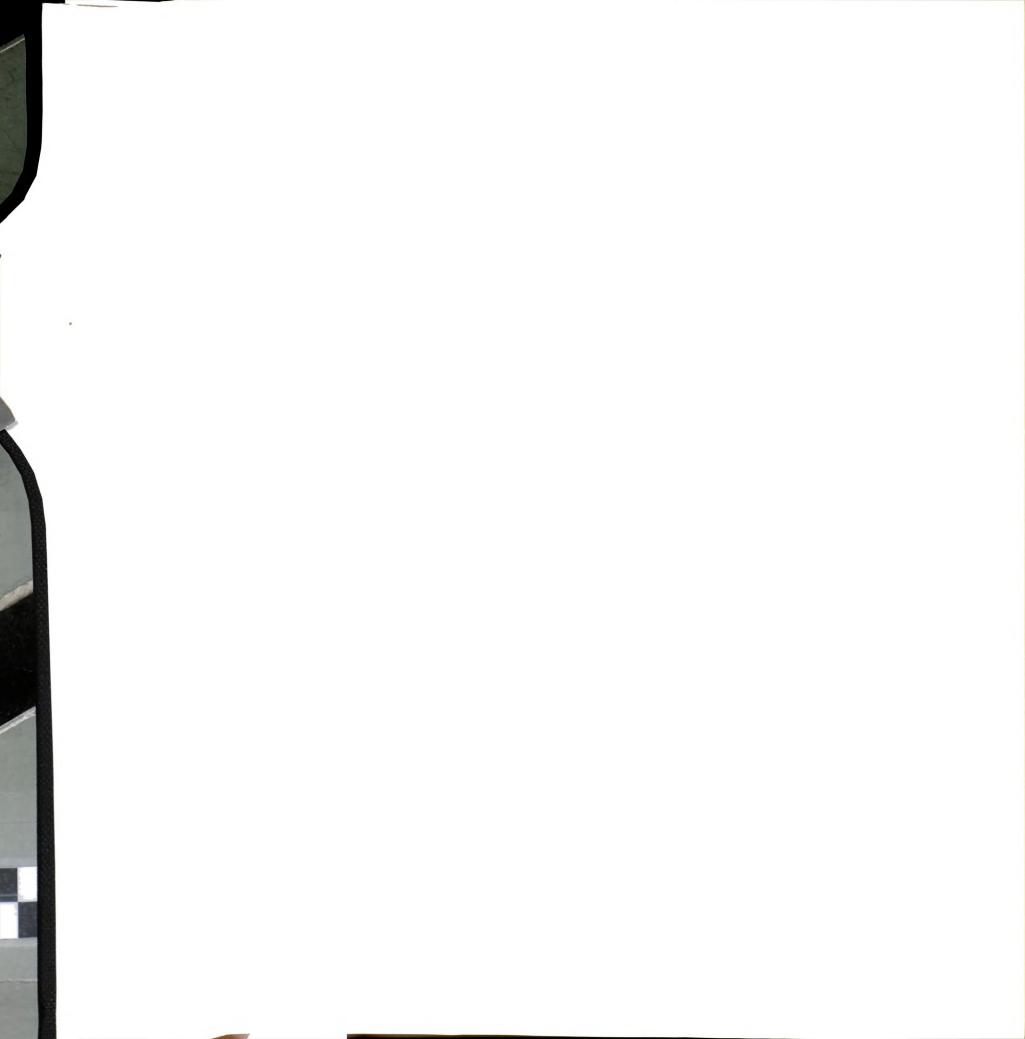
Winter '82 Focal Problems
Student Questionnaire
Page 2

6. Was the number of cases appropriate?

      _____no, should be smaller
      _____yes
      _____no, should be greater

7. Was the amount of biomedical content appropriate?

      _____no, should be more
      _____yes
      _____no, should be less

8. Please rank the cases in order of how stimulating you found them.

      _____Case I (Anoxic Encephalopathy)
      _____Case II (Karen Quinlan)
      _____Case III (Donald C.)
      _____Case IV (Hospital Admission for Dementia)

9. Was the difficulty of the exams appropriate?

      _____no, too difficult
      _____yes
      _____no, too easy

10. Rank the following means of evaluation in order of your preference
    for this course.

      _____objective tests
      _____short answer
      _____in-class essay exams
      _____short take-home papers (3-5 pages each)

11. Any additional comments to help improve this course (use reverse
    side if necessary).

    _____

    _____

    _____

    _____

    _____

FOCAL PROBLEMS – POST TEST

I.   Student #_____

II. Instructions:

   1. Detach the page headed VIEWPOINTS, DEFINITIONS AND ETHICAL
      THEORIES. This will be referred to during the test, and will
      be collected with the answers after you have finished.

   2. The test is composed of 55 true–false questions. Circle the
      T or F provided to the left of each question to indicate your
      response. Later transfer those responses to a machine–scored
      answer sheet using A for True and B for False.

   EXAMPLE:

   A stethoscope is a device for

   T F 1.   listening to the heartbeat
   T F 2.   listening to the lungs
   T F 3.   looking in the ears

      NOTE: In this example there are three separate questions
              beginning with, "A stethoscope is a device for".
              Each way of completing the sentence, 1–3, is a
              separate question requiring a separate response.

   3. Read Scenario I until you come to STOP. Respond to the first
      set of questions. Read Scenario II until you come to a second
      STOP. Respond to the second set of questions. Resume reading
      Scenario II until you come to a third STOP. Respond to the
      third set of questions. In responding to the 3 sets of
      questions, PAY CAREFUL ATTENTION TO THE DIFFERENT INFORMATION
      CORRESPONDING TO EACH SET.

SCENARIO I

Henry Black, a 56 yr old white male, was brought to the emergency room at a San Francisco hospital at 3:00 P.M. one Saturday afternoon, apparently the victim of a massive myocardial infarction (heart attack). His wife, who rode with him in the ambulance to the emergency room, reported that she had returned home from shopping and found him lying on the floor next to a table and a lamp he had upset. In her words, "He was not breathing and was blue." She immediately called an ambulance and began administering mouth to mouth resuscitation. When the ambulance arrived the attendants discovered Mr. Black's heart was beating and immediately placed him on a respirator.

The diagnosis of a massive myocardial infarction was confirmed 36 hours later in the cardiac care unit. Apparently Mr. Black had been without oxygen for a significant length of time; testing showed that he satisfied all the Harvard Criteria for brain death including a flat EEG. The tests had been applied 24 hours apart; the last administration was 48 hours after Mr. Black's admission. A discussion among the staff ensued. Three viewpoints were represented.

STOP

Refer to the page headed VIEWPOINTS, DEFINITIONS, and ETHICAL THEORIES as necessary. RESPOND TO QUESTIONS 1-36.

If you accept VIEWPOINT I, then you must accept

T F 1. DEFINITION A
T F 2. ETHICAL THEORY 1

If you accept VIEWPOINT II, then you must accept

T F   3. DEFINITION B
T F   4. ETHICAL THEORY 2

If you accept VIEWPOINT III, then you must accept

T F   5. DEFINITION C
T F   6. ETHICAL THEORY 3

If you accept DEFINITION A, then you must accept

T F   7. VIEWPOINT I
T F   8. ETHICAL THEORY 1

If you accept DEFINITION B, then you must accept

T F   9. VIEWPOINT II
T F 10. ETHICAL THEORY 2

If you accept DEFINITION C, then you must accept

T F 11. VIEWPOINT III
T F 12. ETHICAL THEORY 3

Assuming you accept DEFINITION A, deciding whether the patient is dead is

T F 13. an ethical issue
T F 14. a factual (medical/technical) issue
T F 15. an issue of how to distribute limited medical resources
T F 16. a legal issue

The claim that the quality of life is more important than the quantity of life in this situation is an assumption of

T F 17. VIEWPOINT I
T F 18. VIEWPOINT II
T F 19. VIEWPOINT III

A distinction between an oxygenated human organism and a living person is an assumption of

T F 20. VIEWPOINT I
T F 21. VIEWPOINT II
T F 22. VIEWPOINT III

The claim, "The cost of maintaining the patient in his present condition relative to other potential uses of the medical resources expended is something which should be taken into account", must be denied by a consistent supporter of

T F 23. VIEWPOINT I
T F 24. VIEWPOINT II
T F 25. VIEWPOINT III

The question, "Would you bury the patient RIGHT NOW?" is a way of objecting to

T F 26. DEFINITION A
T F 27. DEFINITION B
T F 28. DEFINITION C

The claim that DEFINITION A is satisfied in this situation could reasonably be denied by a supporter of

T F 29. VIEWPOINT II
T F 30. VIEWPOINT III

The remark, "Another patient in need of kidneys would probably benefit more from Mr. Black's kidneys than he can, but we can't go around making decisions solely or primarily on those grounds," is a way of objecting to

T F 31. ETHICAL THEORY 1
T F 32. ETHICAL THEORY 2
T F 33. ETHICAL THEORY 3

A person who remarks in this situation, "What's being considered here is euthanasia and I want no part of it," must reject

T F 34. DEFINITION A
T F 35. DEFINITION B
T F 36. DEFINITION C

SCENARIO II

Exactly the same sequence of events occurred as described in SCENARIO I up to the point of the second administration of the tests for the Harvard Criteria. In SCENARIO II Mr. Black began breathing spontaneously before the tests were applied a second time. His condition was otherwise the same. That is, except for the requirement of no spontaneous breathing, he satisfied all of the Harvard Criteria including a flat EEG.

STOP

Refer to the page headed VIEWPOINTS, DEFINITIONS
and ETHICAL THEORIES as necessary. RESPOND TO
QUESTIONS 38-43.

The facts by themselves in SCENARIO II rule out

T F 38. VIEWPOINT I
T F 39. VIEWPOINT II
T F 40. VIEWPOINT II

Suppose someone wants to defend VIEWPOINT II for SCENARIO II. If so,
they must reject

T F 41. DEFINITION A
T F 42. DEFINITION B
T F 43. DEFINITION C

RESUME SCENARIO II

Mr. Black could possibly be maintained for months provided
appropriate measures were undertaken, for instance, tube-feeding or
preventive antibiotic therapy. But, as it turned out, Mr. Black had
executed a "living will". (This is a document which communicates the
wishes of a person pertaining to medical treatment in the event they
are rendered incompetent and unable to express their wishes. Such
documents are legally binding in California where the case of Mr. Black
took place.) There were two wishes expressed by Mr. Black in this
document which are especially relevant to the present situation: (1)
"In the event that I become incapable of living a genuine human
existence and am unable to express my wishes, I DEMAND that no extra-
ordinary medical measures be used to maintain me." (2) "In the event
that the conditions set out in the first clause of (1) are satisfied,
I give my permission to the appropriate medical authorities to

immediately remove any and all of my bodily organs for transplant or research purposes."

<div align="center">STOP</div>

> Refer to the page headed VIEWPOINTS, DEFINITIONS
> and ETHICAL THEORIES as necessary. RESPOND TO
> QUESTIONS 44-55.

In order to remove <u>reasonable doubt</u> about how to carry out Mr. Black's wishes as expressed in his "living will", you would have to clarify the meaning of which of these expressions?

T F 44. genuine human existence
T F 45. unable to express my wishes
T F 46. I DEMAND
T F 47. extraordinary medical measures
T F 48. immediately remove
T F 49. transplant or research purposes

Mr. Black's "living will" provides him with a means of

T F 50. exercising his right to refuse medical treatment
T F 51. preserving limited medical resources
T F 52. exercising his right of informed consent

<u>Assuming</u> Mr. Black's "living will" should be interpreted to mean that he would not want to be treated under the circumstances described, a person who sought to treat him anyway would probably accept

T F 53. VIEWPOINT III
T F 54. DEFINITION C
T F 55. ETHICAL THEORY 2

VIEWPOINTS, DEFINITIONS and ETHICAL THEORIES

This page is provided for reference purposes. The arrangement of the VIEWPOINTS, DEFINITIONS and ETHICAL THEORIES has no significance. VIEWPOINT I is not necessarily associated with DEFINITION A and ETHICAL THEORY 1; VIEWPOINT II is not necessarily associated with DEFINITION B and ETHICAL THEORY 2; nor is VIEWPOINT III necessarily associated with DEFINITION C and ETHICAL THEORY 3.

VIEWPOINTS

I

This patient is dead since he meets DEFINITION A (THE HARVARD CRITERIA). Therefore, we should go ahead and remove his kidneys. There is not a moment to spare if they are to be maintained in good condition for transplant.

II

The patient is not dead because he is breathing and his heart is beating. True, he is in an irreversible coma and will never regain consciousness. In fact, he has reached a point where he should be allowed to die with dignity. After we allow him to die, we can take his kidneys.

III

The patient is not dead. Moreover, we must do whatever is medically possible to keep him alive. Only if he dies despite our utmost efforts can we take his kidneys.

DEFINITIONS

A

Persons are dead when they are: unreceptive and unresponsive, exhibit no voluntary movement, no spontaneous breathing, no reflexes, and have a flat EEG. (THE HARVARD CRITERIA)

B

Persons are dead when respiration and heartbeat cease despite efforts, including mechanical, to maintain them.

C

Persons are dead when they no longer respond to stimulation, are incapable of voluntary movement, exhibit no signs of cognitive activity, and are unable to undertake plans regarding how to conduct their life.

ETHICAL THEORIES

1

Actions are morally correct which result in maximizing the general good, that is, which result in creating the greatest benefit for the greatest number of people. The rights of individuals and their duties derive from and are secondary to maximizing the general good.

2

Actions are morally correct which respect the rights of all individuals concerned. Duties derive from and are secondary to rights. Maximizing the general good is allowable to the extent that it does not violate the rights of individuals.

3

Actions are morally correct which are in accordance with duties. Rights derive from and are secondary to duties. Maximizing the general good is allowable to the extent that it does not involve the violation of duties.

INTERVIEW PROTOCOL
(PRECEPTORS)

I. Quality of Materials

    A. Cases

        1. #

        2. quality

    B. Study questions

    C. Leader's guide

    D. Suggestions

II. Quality of faculty development and suggestions

III. Students in Discussion

    a. Performance and understanding of aims

    B. Interest and enthusiasm

    C. Problem areas

    D. Suggestions

IV. Yourself in discussion   initial - later reactions

    A. Clarity of aims

    B. Groping at times?

    C. Will you improve with experience?

    D. Interest in precepting in future

V. Your view of medical ethics

    A. Importance

    B. "Teachability"

    C. What should it include?

    D. Evaluation of students

VI. Any additional comments

    Lecture scheduling

APPENDIX C

FOCAL PROBLEMS (1983) INSTRUMENTS

FORM A

I Multiple Choice

1. Which of the following is <u>least</u> relevant where treatment de-
    cisions involving "newly" incompetent adults (e.g., Karen
    Quinlan) are concerned?

    a. the patient's prognosis
    b. what the patient would want if competent
    c. the values of the physicians involved
    d. the law

2. In a random clinical trial of two drugs the treatment the
    patient undergoes is chosen by

    a. the patient
    b. the physician
    c. the patient and the physician jointly
    d. none of the above

3. The right of a research subject to informed consent requires

    a. informing subjects of all alternatives
    b. informing subjects of the risks and benefits of the
        research procedures
    c. informing subjects that they may withdraw at any time
    d. all of the above

4. Which of the following is the <u>least</u> relevant to a decision of
    whether to continue a respirator for a terminally ill patient?

    a. quality of life
    b. what the patient wants
    c. what the family wants
    d. the right to life

5. A patient's right to accept or refuse life-saving medical
    treatment

    a. depends on the patient's prognosis
    b. is equal to the physician's right to make treatment
        decisions
    c. imposes a check on what physicians can do
    d. all of the above

Focal Problems Track I
Term 3
Pretest

II   True-False

T F   6.   In cases where the prognosis is very poor (e.g., Karen
           Quinlan) a good way to avoid becoming ensnarled in a messy
           legal situation is not to begin treatment in the first place.

T F   7.   A patient's right to refuse treatment does not imply that
           every voluntary refusal of life-saving treatment should be
           honored.

T F   8.   The right to autonomy (self-determination) is included in
           the constitutional right to privacy.

T F   9.   The distinction between ordinary and extraordinary treatment
           could justify stopping the use of a respirator, but could
           not justify stopping the use of antibiotics.

T F 10.   Enrolling a patient in a research protocol is sufficiently
           justified if the physician sincerely believes it is in the
           patient's best interests.

T F 11.   The right to life implies that life-saving medical treatment
           may never be withheld.

T F 12.   If it is justifiable to start a patient on a respirator, it
           is not justifiable to subsequently stop the respirator (unless
           the patient is or becomes legally dead).

III   CASE DESCRIPTION

     On December 14, 1982, Charlie Brooks Jr. was the first U.S.
prisoner ever executed by lethal injection. In order for the execution
to take place, a physician, Ralph Gray, inspected the arm of Brooks
to insure that a catheter could be inserted into the veins, instructed
a technician on where and how to make the insertion, and stood by as
sodium thiopental, pancuronium bromide and potassium chloride flowed
into Brooks' arm in sequence. He died within minutes.

     As a result of this execution, the question of whether physicians
can be involved in these kinds of executions without violating their
professional obligation to preserve life has been raised. Below are
several claims made in support of Dr. Gray.

     Dr. Gray loaded the pistol, but he did not pull the trigger.
     There was no doctor involved in the actual process of the execution.
     Looking for veins doesn't count.

                                   From Time, Dec. 20, 1982

Focal Problems Track I
Term 3
Pretest

1. Dr. Gray ___should ___should not have participated in the
    execution. Check off and defend your answer in a paragraph
    or two.

Focal Problems Track I
Term 3
Pretest

2. The claims made in support of Dr. Gray ___were ___were not
    satisfactory. Check off and defend your answer in a paragraph
    or two.

Focal Problems Track I   Student #_____
Term 3   Preceptors_____
Pretest

FORM B

I   Multiple Choice

1. Which of the following is <u>not</u> necessary to justify a research
    protocol?

    a. informed consent of the subjects
    b. benefit to the subjects
    c. benefit to future patients
    d. a favorable overall risk/benefit ratio

2. Which of the following is <u>least</u> important to adhering to a
    patient's refusal of life-saving treatment?

    a. patient competence
    b. an informed patient
    c. physician agreement
    d. a-c are equally important

3. In a random clinical trial of two drugs, the treatment the
    patient undergoes is chosen by

    a. the patient
    b. the physician
    c. the patient and the physician jointly
    d. none of the above

4. Which of the following is <u>least</u> relevant where treatment
    decisions involving "newly" incompetent patients (e.g., Karen
    Quinlan) are concerned?

    a. the patient's prognosis
    b. what the patient would want if competent
    c. the values of the physicians involved
    d. the law

5. Which of the following is the strongest justification for
    honoring a refusal of treatment?

    a. the agreement of the family
    b. the patient's rights
    c. the agreement of the physician
    d. all of the above are equally strong

II True-False

T F 6. If the life of persons is infinitely valuable, then with-
         holding life-saving medical treatment is never justified.

Focal Problems Track I
Term 3
Pretest

T F 7. Patient autonomy and "death with dignity" require that terminally ill patients never be treated against their expressed wishes.

T F 8. The decision of whether to start a patient on a respirator should <u>not</u> be based on whether there may be reasons to stop it later.

T F 9. Active euthanasia (e.g., giving a lethal injection) is never ethical and passive enthanasia (e.g., "no codes") is sometimes ethical <u>because</u> killing is always morally worse than allowing to die.

T F 10. The use of an experimental treatment is sufficiently justified if there is a possible benefit to future patients and risks to the subjects are minimal.

T F 11. A person's right to life excludes the use of quality of life considerations in a decision to withhold or withdraw treatment.

T F 12. The decision of the New Jersey Supreme Court in the case of Karen Quinlan illustrates the tendency of the courts to leave the standards for practice of medicine to the medical profession.

III   CASE DESCRIPTION

On December 17, 1976, a baby boy was born $15\frac{1}{2}$ weeks prematurely and weighing 514 grams. He suffered from numerous afflictions, many of which were iatrogenic. On December 24 he was placed on a respirator. (One of the reasons given by a doctor for respirator dependence was that it "hurts like hell every time he breathes." The infant had developed rickets resulting in numerous broken ribs.) The prognosis was bleak indeed; survival was unprecedented.

The parents had objected to the decision of Dr. Farrell, the attending physician, to ventilate the infant. They preferred to allow him to die. They described Dr. Farrell's response to their objection as follows:

When we objected to the decision Dr. Farrell accused us of wanting to "play God" and to "go back to the law of the jungle"..."I would not presume," he told us, "to tell my auto mechanic how to fix my car."

From the <u>Atlantic</u>, July 1979

Focal Problems Track I
Term 3
Pretest

1. The infant ___should ___should not have been ventilated.
Check off and defend your answer in a paragraph or two.

Focal Problems Track I
Term 3
Pretest

2. Dr. Farrell's responses to the parents' objections ___were
    ___were not satisfactory. Check off and defend your answer in
    a paragraph or two.

HM 214   STUDENT QUESTIONNAIRE   WINTER 1983

1. Preceptors' Names_____

2. How much will your experience in this course help you deal with
    ethical problems in your practice?

    _____a great deal

    _____somewhat

    _____very little

    _____don't know

    Comments:_____

    _____

3. Please rate the following course areas.

	low 1	2	average 3	4	high 5
usefulness of lectures	___	___	___	___	___
interest in lectures	___	___	___	___	___
usefulness of discussion groups	___	___	___	___	___
interest in discussion groups	___	___	___	___	___
skill of preceptors	___	___	___	___	___
quality of readings	___	___	___	___	___
quality of discussion guides	___	___	___	___	___
quality of examinations	___	___	___	___	___

4. Was the number of cases appropriate?

    _____no, should be greater

    _____yes

    _____no, should be fewer

HM 214
Student Questionnaire
Winter 1983

5. Was the amount of biomedical content appropriate?

    _____no, should be less

    _____yes

    _____no, should be more

6. Please rank the cases in order of overall interest and importance.

    _____Case I (anoxic child)

    _____Case II (Karen Quinlan)

    _____Case III (Donald C.)

    _____Case IV (man with MS)

    _____Case V (leukemic woman)

7. Any additional comments to help improve this course: