

Fall 1978

GAINS IN MORAL REASONING THROUGH  
TWO DISCUSSION METHODS UNDER  
TEACHING ASSISTANTS VARYING IN  
SOCRATIC SKILL

LEON SWARTZENDRUBER  
*University of New Hampshire, Durham*

Follow this and additional works at: <https://scholars.unh.edu/dissertation>

---

**Recommended Citation**

SWARTZENDRUBER, LEON, "GAINS IN MORAL REASONING THROUGH TWO DISCUSSION METHODS UNDER TEACHING ASSISTANTS VARYING IN SOCRACTIC SKILL" (1978). *Doctoral Dissertations*. 2336.  
<https://scholars.unh.edu/dissertation/2336>

This Dissertation is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact [nicole.hentz@unh.edu](mailto:nicole.hentz@unh.edu).

## INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.
2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.
3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.
4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.
5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.

University  
Microfilms  
International

300 N. ZEEB ROAD, ANN ARBOR, MI 48106  
18 BEDFORD ROW, LONDON WC1R 4EJ, ENGLAND

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark .

1. Glossy photographs \_\_\_\_\_
2. Colored illustrations \_\_\_\_\_
3. Photographs with dark background \_\_\_\_\_
4. Illustrations are poor copy \_\_\_\_\_
5. Print shows through as there is text on both sides of page \_\_\_\_\_
6. Indistinct, broken or small print on several pages  throughout  
\_\_\_\_\_
7. Tightly bound copy with print lost in spine \_\_\_\_\_
8. Computer printout pages with indistinct print \_\_\_\_\_
9. Page(s) \_\_\_\_\_ lacking when material received, and not available  
from school or author \_\_\_\_\_
10. Page(s) \_\_\_\_\_ seem to be missing in numbering only as text  
follows \_\_\_\_\_
11. Poor carbon copy \_\_\_\_\_
12. Not original copy, several pages with blurred type \_\_\_\_\_
13. Appendix pages are poor copy \_\_\_\_\_
14. Original copy with light type \_\_\_\_\_
15. Curling and wrinkled pages \_\_\_\_\_
16. Other \_\_\_\_\_

7909430

SWARTZENDRUBER, LEON  
GAINS IN MORAL REASONING THROUGH TWO  
DISCUSSION METHODS UNDER TEACHING ASSISTANTS  
VARYING IN SOCRATIC SKILL.

UNIVERSITY OF NEW HAMPSHIRE, PH.D., 1978

University  
Microfilms  
International 300 N. ZEEB ROAD, ANN ARBOR, MI 48106

GAINS IN MORAL REASONING THROUGH TWO DISCUSSION  
METHODS UNDER TEACHING ASSISTANTS VARYING  
IN SOCRATIC SKILL

BY

LEON SWARTZENDRUBER

B.A., Whittier College, 1963

M.A., University of Illinois, 1966

A DISSERTATION

Submitted to the University of New Hampshire  
in Partial Fulfillment of  
the Requirements for the Degree of

Doctor of Philosophy

in

Psychology

September, 1978

This dissertation has been examined and approved.

*Lance K Canon*

\_\_\_\_\_  
Dissertation director, Lance K. Canon  
Associate Professor of Psychology

*John G. Chaltas*

\_\_\_\_\_  
John G. Chaltas, Associate Professor  
of Education

*David E. Leary*

\_\_\_\_\_  
David E. Leary, Assistant Professor  
of Psychology

*Niel B. Lubow*

\_\_\_\_\_  
Niel B. Lubow, Assistant Professor  
of Philosophy

*David I. Schickedanz*

\_\_\_\_\_  
David I. Schickedanz, Assistant  
Professor of Psychology

*8/11/78*

\_\_\_\_\_  
Date

## ACKNOWLEDGMENTS

It was Carl Rogers who said, "I have always had the feeling that if I was given some opportunity to do the thing I was most interested in doing, everything else would somehow take care of itself." The process, not necessarily the end result, of doing this dissertation has helped to confirm Rogers' wisdom for me in a personal sense. The key here was having the opportunity to do what I felt committed to do.

That opportunity was generously extended to me by my entire committee, but most notably by Dr. Canon, who spent many, many hours in consultation with me about the progress of the study. Also, a different type of assistance was provided by my work-study assistant, Ms. Patricia M. Miller, to whom I am indebted for her extraordinarily conscientious, efficient, accurate, and even insightful analysis of much, much data.

A debt of a different kind is owed to the teaching assistants, who were the cadre or backbone of this study. In addition, the scores of introductory psychology students whose cooperation was essential are to be thanked.

I thank Dr. James Rest for granting me permission to reproduce the Heinz dilemma from the copyrighted Defining Issues Test.

Appreciation is due to Ms. Trixie McLean, who typed the manuscript. She was most accommodating under difficult conditions.

Quite helpful was the partial funding of this research by the Graduate School and the Research Office, who both supplied me with Central University Research Fund Grant S-92. Also, the generous use of their photocopying equipment came in quite handy.

Finally, my deepest debt is owed to my beloved wife, Mary Lee, who with my equally beloved, very young daughter, Kristen, supported me and thus the dissertation in many, many ways.



## TABLE OF CONTENTS

LIST OF TABLES.....	vii
ABSTRACT.....	viii
I. INTRODUCTION.....	1
From Knowing to Thinking to Moral Reasoning.....	2
Moral Reasoning: Advances by Kohlberg.....	5
Socrates, Kohlberg, and their critics.....	5
Kohlberg's 5-stage model.....	7
More of Kohlberg's critics.....	11
Rest to the Rescue: The Defining Issues Test.....	12
Kohlberg, Rest, and Moral Discussions in Groups...	15
II. METHOD.....	19
The Original Plan.....	19
The Revised Plan.....	20
Teaching Assistants.....	21
Selection.....	21
Training and supervision.....	22
Experimental Design.....	24
Procedure.....	25
Method K(A): Discussion of moral dilemmas...	25
Method Q(B): Discussion of general issues...	30
Subjects.....	32
III. RESULTS.....	33
Measures Recorded.....	33
Primary variables: Mostly experimental.....	33
Secondary variables: Mostly exploratory.....	37
Values and Relations Found.....	40
Experimental findings.....	40
Exploratory findings.....	43
IV. DISCUSSION.....	47
Main Hypotheses and Results: A Rundown.....	47
Conditional Changes on DIT Scores.....	48

Dogmatism and DIT Scores.....	53
Student Satisfaction with Methods.....	54
Exploratory Findings and Implications.....	56
Student ratings.....	56
TA ratings of students.....	58
The essay question on loving.....	59
Some Final Words: Early Steps.....	60
REFERENCES.....	63
APPENDIX A. THE ORIGINAL SIX-STAGE COGNITIVE- DEVELOPMENTAL MODEL OF MORAL REASONING PROPOSED BY KOHLBERG.....	67
APPENDIX B. COURSE SYLLABUS: A CHARTER FOR SKIPPER, CADRE, AND CREW.....	70
APPENDIX C. GROUP DISCUSSIONS: METHOD A.....	81
APPENDIX D. SOCIAL ISSUES FOR GROUP DISCUSSIONS UNDER METHOD A: THE FIRESIDE FIFTY..	83
APPENDIX E. GROUP DISCUSSIONS: METHOD B.....	89
APPENDIX F. EXEMPLARY DISCUSSION QUESTIONS WRITTEN BY STUDENTS UNDER METHOD B...	91

LIST OF TABLES

1. One of the Six Moral Dilemma Exercises on the Defining Issues Test.....	14
2. Example of a Moral Dilemma and Probe Questions Based on <u>Walden Two</u> .....	28
3. Student Evaluation of Discussion: The Discussion Groups.....	38
4. Correlations Between Selected Items Rated by Students Who Used the K-Q Sequence of Methods.....	44

## ABSTRACT

### GAINS IN MORAL REASONING THROUGH TWO DISCUSSION METHODS UNDER TEACHING ASSISTANTS VARYING IN SOCRATIC SKILL

by

LEON SWARTZENDRUBER

Moral growth as part of psychological growth was a goal in an introductory psychology class. Meeting often were 24 six-student discussion groups, complementing the lectures. Half of the groups ran six weeks on one of two methods, then ran six more weeks on the other; this sequence was reversed in the other groups. The two sequences, not the two methods, were compared. As discussion leaders, teaching assistants (TAs) each led two groups, one on each sequence. On Socratic skill, that is, talent in promoting discussion, TAs clearly differed. Thus, sequences and skill levels were completely crossed.

The two discussion methods were called method K, after Kohlberg, and method Q, after the questions it featured. Under method K, adapted from Galbraith and Jones, each TA would read aloud a moral dilemma about an assigned paperback, such as Walden Two. Students would then air their views and reasons fully. Next the TA would summarize the emerging points. Finally, she would announce a Key Person for the session, that is, a group member whose prior remarks, especially his reasons,

were to be briefly written up as homework by everyone else. Thus, students' ears were put to work.

Under method Q, in contrast, homework would come first. After reading the paperback assignment, each student would write a brief discussion question on it. Questions were supposed to integrate the paperback with the course text or with life experience. They were also to be controversial, but not necessarily moral. At the group meeting each student would take turns in using her question to get a short discussion going, the TA providing direction as needed.

The Defining Issues Test (DIT), devised by Rest, assesses moral reasoning. On it a student selects and weights certain reasons useful in resolving certain moral dilemmas. The test yields a weighted percent of principled (Kohlberg's postconventional) moral reasoning (P%). A pretest and a posttest at the beginning and the end of the course rendered DIT means of 43.03 P% and 44.20 P% for the students on sequence K-Q; means were 42.27 P% and 45.53 P% for sequence Q-K. Thus, the mean gains for sequences K-Q and Q-K were 1.17 P% and 3.26 P%, but the difference between the two gains was not significant.

In weekly TA meetings, where TAs practiced discussing dilemmas among themselves, I rated the TAs' Socratic skills. By then pairing each TA's rating with the mean P% gain for his group, no significant correlation emerged across the K-Q groups, but a significant one of .65 did obtain in the Q-K ones. Reasons are given, however, to view this relation with caution.

On a scale from 1 to 5 (awful to excellent) students rated the educational value of aspects of the course. In the

early weeks, methods K and Q were rated at means of 3.98 and 4.22, significantly different. In the final week, sequences K-Q and Q-K as combinations were awarded means of 4.16 and 4.58, again significantly different.

The Dogmatism Scale, developed by Rokeach, was completed in the first week. It correlated significantly with the DIT pretest and the posttest at  $-.21$  and  $-.27$ , but did not correlate significantly with DIT gain. Nor did it help in various analyses of covariance.

An interpretation of general results focuses mainly upon the Q-K sequence and Socratic skills. Under method Q students tailor questions to their own interests and share the assignment of discussing them with others. Soon such sessions build involvement and rapport, thus setting the stage for method K, directed squarely at moral reasoning. With any stage fright now down among players, the director or TA can be as effective as her Socratic skill permits. With it she can adroitly and tactfully pit one view against another to foster creative conflict within each student. If nondefensive, he can grow from it by searching for new grounds, hopefully postconventional. But the search never ends, as moral philosophers, an argumentative lot, agree.

## CHAPTER I

### INTRODUCTION

Together the two monstrous bombs brought instant death to more than one hundred thousand in August of 1945. When news of this broke, the ultimate button pusher behind it all, President Truman, was both denounced and praised throughout the world in the strongest of terms. Whether warranted or not, his decision to bomb Hiroshima and Nagasaki he saw as morally just. Although two heavily populated cities were to be leveled and atomically poisoned, this cost in human suffering--so he and his advisors reasoned--would be less than the misery of invading and cleaning out the Japanese mainland with legions of war weary troops, who would slaughter countless others and be slaughtered themselves.

On the lighter side, it is of interest that the President, whether defensively or not, usually displayed a maxim from Mark Twain on his desk in the Oval Office: "Always do right. This will gratify some people and astonish the rest."

A necessary but not always sufficient condition for doing right in a given situation is a comprehensive knowledge of all relevant considerations and a great deal of critical thinking about them. These two key aspects in the minds of decision makers and the voters who elect them can be promoted to some extent through education. Probably the most helpful thorough analysis of knowledge and critical thinking is contained in the landmark work of Bloom and associates (1956). In this

classic of education they give their definitions of knowing and thinking, the latter of which is what I mean by critical thinking:

Knowledge...involves the recall of specifics and universals, the recall of methods and processes, or the recall of a pattern, structure, or setting....  
The recall situation involves little more than bringing to mind the appropriate material. (p. 201; emphasis added)

Thinking or abilities and skills refer to organized modes of operation and generalized techniques for dealing with materials and problems. The materials and problems may be of such a nature that little... information is required...and can be assumed to be part of the individuals' general fund of knowledge. Other problems may require specialized and technical information at a rather high level such that specific knowledge and skill in dealing with the problem...are required. (p. 204; emphasis added)

From the emphasized words in these definitions it can be seen that knowing and thinking are hierarchical, the first being necessary but not sufficient for the second. Incidentally, it should also be noted for the sake of completeness that neither of these cognitive goals are completely separable from affective ones, as articulated by Bloom's colleagues, Krathwohl and associates (1964).

#### From Knowing to Thinking to Moral Reasoning

While knowing would seem less important than thinking as goals of education, this priority would not be obvious at all to the proverbial extraterrestrial visitor assigned to observe a random sample of educators in classrooms. Here the observer from outer space could easily find more assignments and tests designed to promote knowing than those aimed at thinking. A likely reason for this apparently upside down



practice is that the first is easily and clearly attainable, whereas the second is not. In fact, the teacher who shoots for the second may be in for a lot of grief. This seemingly masochistically inflicted condition is endurable, however, if one eye is always kept on the long-term backdrop of deeper gratifications that often ensue.

Although the reasons for the reluctance to venture beyond the knowledge goal are probably too complex and numerous to detail here, two of the most salient in my own experience are these. First, too much classroom experience has already reinforced too many students too much for merely storing information without reshuffling it. Thus, many students merely equate education with information storage. Implicit here is the notion that out in the real world the thinking process will automatically use the stored information without the inconvenience of prior practice. Study habits are then set accordingly.

Second, there are both outside and mutual pressures upon the student and the teacher to evaluate each other. Since the evaluations often have heavy consequences--whether used by employers in hiring or deans in granting tenure--fairness in making them becomes a weighty concern. Fairness is reasonably clearcut in evaluating the knowing goal, thereby encouraging student acceptance. However, fairness can easily get murky in evaluating the thinking goal, thus disturbing some students. Consider the one who anonymously wrote this evaluation of four lengthy multiple choice exams, whose questions were aimed at thinking more often than at knowing (Swartzendruber, 1974).

Exam[s] were good but at times tricky. If "cases" of using the materials we've received [of applying what we have read] are used on the tests, why aren't they discussed in lectures?

This student seems to be saying that the principles (materials) that have been taught (received) sometimes appear on the tests in connection with new instances or "cases" that have not already been discussed in lectures, and this practice is tricky and unfair; it requires critical thinking.

Student satisfaction, though important, is not the whole picture. An important question still remains, To what extent can educators improve thinking skills? Although broad measures of thinking skills do exist, such as the Cornell Critical Thinking Test (Ennis and Millman, 1971) or the Critical Thinking Appraisal (Watson and Glaser, 1964), it may be too much to expect broad, overall gains from a single course, such as introductory psychology. Instead, only one aspect of thinking or reasoning might be selected as a more reasonable goal to work on.

Most of the courses that I have taught try to emphasize thinking. This is particularly true in introductory psychology, where my students regularly meet in small discussion groups to think and talk about social issues that are controversial. Many of these issues are moral.

Perhaps, then, a more modest goal than overall gains in general critical thinking skills would be gains in a particular kind of reasoning, say, moral reasoning. Indeed, in a course such as mine that emphasizes social psychology the study of morality or cooperation among people fits right in.

In agreement with this premise would be one of the strongest figures in the history of social psychology, William McDougall (1908), who wrote, "The fundamental problem of social psychology is the moralization of the individual by society." (p. 6)

#### Moral Reasoning: Advances by Kohlberg

Few informed observers would deny that the most influential leader in the field of moral education is Harvard's Lawrence Kohlberg. Yet this is not to say that he and those like him have not been the objects of much criticism.

#### Socrates, Kohlberg, and their critics

Little in the realm of psychological ideas is new. That is why both Socrates and Kohlberg, although separated by several millenia, reached essentially the same core of conclusions. Consider the common set. First, virtue or justice has only one ideal form, regardless of the circumstances surrounding it. Second, truly understanding what virtue or justice is enables one to select it over all other contenders. Third, virtue or justice is known by intuition, not by convention or taking a poll of what others think or feel.

Perhaps the unacceptability of these idealistic assertions to the majority of Socrates' fellow Athenians was partly responsible for his execution by hemlock. Although Kohlberg is unlikely to meet the same fate, he has quite probably stirred up more controversy than anyone else in the field of moral education. For example, consider your own reaction to the following question posed by Kohlberg (1971a):

Is it so surprising that psychologists have never believed or understood Socrates? It is hard to understand if you are not stage 6. (p. 232)

Obviously the zealous confidence--some would say over-confidence--behind these words or between the lines has sparked vigorous criticism from a variety of psychologists and moral philosophers. Most controversial of his concepts has been stage 6, the alleged ideal form of virtue or justice. The charge of elitism has been raised, as typified in this rebuke by Fraenkel (1976):

Only three individuals have been identified as being at stage 6--Kohlberg himself, one of his graduate students, and Martin Luther King. (p. 219)

In the same breath Fraenkel voices a related criticism:

The fact that there really are six stages has by no means been established. (p. 219)

Kohlberg, however, can give with the punches. Although the preceding critical quotations have only scratched the surface of a much larger body of objections, they show what Kohlberg has had to address. Accordingly, he has subsequently modified his position to eliminate stage 6 from all but purely theoretical considerations. Thus, what remains, for practical purposes, is a 5-stage model, as shown in a recent manual for scoring verbal responses (Kohlberg and associates, 1976).

Incidentally, when I once told him that my own inability to understand stage 6 did not persuade me that, thus, no one else could grasp it either, he responded that he was not too sure that he himself understood it!

### Kohlberg's 5-stage cognitive-developmental model

Briefly, Kohlberg is saying that all moral decisions are based to some extent on underlying moral reasoning. This reasoning for any one thinker at any one period of life is pervasive and all encompassing as a complete mode of solving all moral problems. Just as younger and older children use different modes of reasoning to solve the Piagetan problem of conservation of volume, so do younger and older children use different modes of reasoning to solve moral problems. Whether in the moral or the broader logical realm, children advance from one mode to another by accommodation, not mere assimilation. That is, the underlying cognitive structure changes.

Kohlberg maintains that the developing human gradually passes from one mode to another, from one stage to another. Each stage has universally recognizable characteristics. For any two sequentially adjacent stages, the higher or more advanced one mediates a more satisfactory solution to a broader range of moral problems or dilemmas. In determining how satisfactory a solution is, Kohlberg relies substantially upon a modern formulation of Kant's categorical imperative, namely, Rawl's original position. Rawls (1971) maintains that to settle a conflict of interest between or among parties, one should assume that he or she could hypothetically be asked to actually be in any of the roles of the parties involved, but that at present there is a "veil of ignorance" that is preventing one from knowing the actual role that he or she would have to carry out. Thus, in such an original position, each player in

the "game" is likely to distribute justice as evenly as possible, rather than possibly winding up on the short end.

As originally conceived, Kohlberg postulated six stages, described completely in Appendix A. It can be ignored at least for now, however, by abstracting from the six a more practical, recent, condensed five-stage model that collapses the old stages 5 and 6. In the present context it will be presented in descending, rather than ascending, order; the top two stages --4 and 5--are far more important in describing the moral reasoning of college students, the experimental subjects in the study to be reported shortly.

Postconventional level--Formerly stages 6 and 5, this level now consists of only the latter. Stage 5, the autonomous stage, consists of principles whose validity rests more upon reciprocity and equality of human rights than upon majority decisions, although public consensus derived through legitimate procedures is still thought to be necessary for social order. This stage requires more individualized thinking--seemingly akin to Bloom's higher cognitive skills--than do any of the lower stages.

Conventional level--This level boils down to conforming to social demands. Stage 4, the law-and-order stage, consists of both the overt and covert rules of the society at large. Stage 3, the good-boy-and-nice-girl stage, consists of the social expectations of people in one's family, group, or most immediate circle of acquaintances. Stage 4 sanctions are more official, whereas stage 3 sanctions are more interpersonal. Both stages require accepting a "bag of virtues"--requiring

only Bloom's lowest category of skills (such as the ability to memorize and accept the ten commandments or the code of the gang).

Preconventional level--This level is egocentric in orientation, focusing only upon the physical or hedonistic results of one's actions. Stage 2, the reciprocity stage, consists of agreements or deals reached by bargaining, in which each party attempts to further only her or his own rewards or interests. Stage 1, the obedience stage, consists of the patterns that one must follow to avoid discomfort, especially as dished out by stronger or more authoritative agents. Stage 2 is to stage 1 as positive is to negative, as approaching is to avoiding, and as reward is to punishment. Both stages are so basic as to be found almost exclusively in young children and--although Kohlberg has not offered this definition--perhaps in animals that have been appetitively or aversively conditioned.

Some comments and findings--Based upon the more-cognitive-than-developmental work of Piaget (1948) and the more-developmental-than-cognitive work of Dewey (1916), the cognitive-developmental model of Kohlberg is characterized most strongly by the stage-sequence notion. Specifically, sequential progression through cognitive stages is necessary but not sufficient for correlated progression through moral stages. For example, as a star student in both college and law school, Richard Nixon's cognitive development was surpassed by few. Yet it outstripped his moral development appreciably, if Kohlberg (1975) is correct in his assessment, "No public or private word or deed of Nixon ever rose above Stage 4." (p. 674)

What is necessary and sufficient for moral development is both the cognitive apparatus and the appropriate environmental conditions, such as education. An attempt to provide the latter will be discussed in great detail in the next chapter.

Progression through the moral stages nearly always proceeds from one stage to the next without skipping or regressing (Turiel, 1966, 1969), except that transition from stage 4 to stage 5 is occasionally marked by a theoretically unparsimonious yet only temporary regression to stage 2 (Kohlberg and Kramer, 1969). Further, this step-by-step sequencing has been observed across a variety of cultures, at least thirteen to date, as diverse as Canada, Israel, India, and Taiwan (Kohlberg, 1971b); yet cleaner, more clearly reported research techniques and still more cultures are needed to satisfy critics such as Simpson (1974) and Kurtines and Greif (1974), all of whom have raised some worthy questions, which will be addressed shortly.

The stages are more cognitive than affective and do not indicate how compassionate, how kind, or how selfless a person may be. Moral stages are not an indication of a person's character. A better measure of the latter may be the extent to which a person lives up to what he or she honestly thinks is right, regardless of what stage the thinking is at.

Finally, the majority of a person's moral judgments are nearly always found at a single stage, but not exclusively. Adjacent stages also account for substantial parts of a person's moral reasoning. Thus, on matters of authority and life a



student may prefer stage-5 arguments, but on a matter such as sex she may be more comfortable with stage 4.

#### More of Kohlberg's critics

Kohlberg's approach is more "soft" than "hard" and more innovative than confirmatory. Thus, it is no surprise that he became a target of two rigorous, empirically oriented young psychologists who were rapidly progressing through the Johns Hopkins doctoral program together. While there they took a close look at the scientific adequacy of Kohlberg's work, particularly his observational methods, and came up with a penetrating report that was published in a highly reputable scientific journal.

The report is by Kurtines and Greif (1974), and its thrust has four parts. First, for Kohlberg the connection between moral behavior and moral reasoning is unspecified and poorly correlated. Second, the Kohlberg model seems somewhat arbitrary, and its associated measuring scale lacks standardization; it has never been published and is continually being revised, thus being generally unavailable for research. Third, the reliability of the scale is unreported and thus unknown. Fourth, construct validation for the scale has been lacking.

These charges are serious, but as will soon be shown, only the first remains largely untouched. Let us at least expand upon it. Moral reasoning is based on moral attitudes, and the poor correlation between attitudes and behavior in a variety of areas is well documented (Calder and Ross, 1973).

Kohlberg and his entire staff are well aware of the problem. An excellent, insightful analysis of it was undertaken by Brown and Herrnstein (1975), who concluded that although moral reasoning is one notable determinant of behavior, there are certainly others of at least equal power. Prominent among them is a person's degree of moral concern, that is, the extent to which a person is concerned about doing what he or she thinks is morally right or just. Brown and Herrnstein call it "character." Still another host of largely nonmoral influences on behavior come from the specific situation in which the action occurs. This view is consistent with the presently predominant concept of person-situation interactionism championed by Mischel (1973).

Regarding the remaining three cogent criticisms of Kurtines and Greif, note that they all deal with the problem of the measuring instrument by which moral reasoning is assessed. However, the two co-workers do concede that perhaps "the stages do reflect actual development and...the general lack of evidence reflects the inadequacy of the measuring device..." (1974, p. 469). Thus, if an adequate measuring device could be constructed, there is a whole new ball game to be played.

Somewhat coincidentally, Kurtines and Greif's report was in press at the same time that another highly related one was. Let us take a look at the heart of the latter.

#### Rest to the Rescue: The Defining Issues Test

In response to the accumulating need for an instrument having reliability, validity, availability, standardization,

and a stable set of norms, Rest and his colleagues (1974a) developed the Defining Issues Test (DIT), as well as a manual for it (Rest, 1974b). The instrument consists of six moral dilemmas, each of which is followed by a list of reasons for choosing one horn or the other, as shown in Table 1. The testee rank orders the reasons in order of their importance so as to provide a weighted index of the extent to which he or she prefers postconventional--also called principled--reasoning. By averaging the amount of principled reasoning across all six dilemmas, a single percentage score, labeled P%, is obtained.

The use of P% is a rather clever way of circumventing the question--voiced by Kurtines and Greif--of whether the 5-stage model is arbitrary. By simply postulating that the last stage, principled reasoning, is different from the first four stages, and then counting the number of principled judgments and weighting them, a continuous scale having useful properties is obtained. Empirical support for the reliability, validity, and usefulness of P% was gathered in an important dissertation by Cooper (1972), who compared P% with other ways of scoring the DIT.

Moreover, Rest and associates (1974) have shown that the DIT addresses Kurtines and Greif's concerns with the following properties. Its test-retest reliability over two weeks is 0.81. Its correlation with Kohlberg's measure is 0.68. Thus, Rest does not regard the DIT as equivalent to Kohlberg's measure but instead considers it an alternate way to assess moral reasoning. The DIT correlates at -0.60 with a Law and Order scale and at 0.63 with a Libertarianism scale,

TABLE 1

ONE OF THE SIX MORAL DILEMMA EXERCISES ON THE DEFINING ISSUES TEST\*  
HEINZ AND THE DRUG

In Europe a woman was near death from a special kind of cancer. There was one drug that the doctors thought might save her. It was a form of radium that a druggist in the same town had recently discovered. The drug was expensive to make, but the druggist was charging ten times what the drug cost to make. He paid \$200 for the radium and charged \$2000 for a small dose of the drug. The sick woman's husband, Heinz, went to everyone he knew to borrow the money, but he could only get together about \$1000, which is half of what it cost. He told the druggist that his wife was dying, and asked him to sell it cheaper or let him pay later. But the druggist said, "No, I discovered the drug and I'm going to make money from it." So Heinz got desperate and began to think about breaking into the man's store to steal the drug for his wife.

Should Heinz steal the drug? (Check one)

Should steal it                       Can't decide                       Should not steal it

IMPORTANCE:

Great	Much	Some	Little	No	
					1. Whether a community's laws are going to be upheld.
					2. Isn't it only natural for a loving husband to care so much for his wife that he'd steal?
					3. Is Heinz willing to risk getting shot as a burglar or going to jail for the chance that stealing the drug might help?
					4. Whether Heinz is a professional wrestler, or has considerable influence with professional wrestlers.
					5. Whether Heinz is stealing for himself or doing this solely to help someone else.
					6. Whether the druggist's rights to his invention have to be respected.
					7. Whether the essence of living is more encompassing than the termination of dying, socially and individually.
					8. What values are going to be the basis for governing how people act towards each other.
					9. Whether the druggist is going to be allowed to hide behind a worthless law which only protects the rich anyhow.
					10. Whether the law in this case is getting in the way of the most basic claim of any member of society.
					11. Whether the druggist deserves to be robbed for being so greedy and cruel.
					12. Would stealing in such a case bring about more total good for the whole society or not.

From the list of questions above, select the four most important:

- Most Important \_\_\_\_\_
- Second Most Important \_\_\_\_\_
- Third Most Important \_\_\_\_\_
- Fourth Most Important \_\_\_\_\_

\* © James Rest, 1972

thus showing some construct validity. Over thousands of subjects at various educational levels, the means for P% are as follows (Rest, 1976): junior high, 22; senior high, 32; college, 42; graduate school, 53; doctoral students in moral philosophy and political science, 65. Since the DIT is an objective inventory, perhaps the mere recognition of phrases having a highly moralistic ring could enable a testee to fake high scores, but McGeorge (1975) tested this question empirically and found that the DIT is not susceptible to faking.

One dissertation committee member suggested that the construct validity of the DIT could be strengthened by the classic multitrait-multimethod matrix method of Campbell and Fiske (1959). Here the trait labeled P% would be measured by the DIT method, the Kohlberg interview method, and perhaps some other methods, all of which should correlate well with the DIT--about as well as it has already been found to correlate with the Kohlberg interview, that is 0.68. Also, the forced choice method of the DIT would be used to measure other traits which should then not correlate as highly with P%. Unfortunately, I am not aware of any comprehensive study that has used the multitrait-multimethod technique to investigate the DIT.

#### Kohlberg, Rest, and Moral Discussions in Groups

Underlying the promotion of moral reasoning are two fundamental propositions (Kohlberg and Turiel, 1971). First, the student must experience conflict--what Festinger calls cognitive dissonance--in applying her stage of reasoning to some problem. (For example, stage-4, law-and-order morality

entails conflicting values in providing a rationale for whether to illegally help slaves escape to Canada through the Underground Railroad before the Civil War.) Second, the student's growth in reasoning is greatly helped if she is exposed to reasoning at the next higher stage. (Continuing with the example, if she hears Rawls' stage-5 "veil of ignorance" applied to the runaway slave case and if she truly comprehends this application, she could reduce the stage-4 conflict by advocating help for the slave on the grounds that she would not want to have to play the role of a runaway slave who got caught. Thus, in this particular area of her moral reasoning she would be growing toward stage 5.)

In discussion groups the two conditions can be easily met. By discussing moral dilemmas in a warm atmosphere where a mixture of stages is present among different students, stages of judgment in junior-high and high-school students have been raised in experimental work by Blatt and Kohlberg (1975).

The value of discussions in promoting moral reasoning is also implied in the following words by Rest (1974b):

As people participate in moral discussion and publicly support and defend certain sides of an issue, they are making public moral judgments which have influencing effects on others.  
(page 5-8)

In fact, Rest (1974a) notes that the discussion of controversial issues is the key feature in Kohlbergian programs, though Rest advocates enriching this strategy with still other methods, such as role playing.

In a doctoral dissertation that Kohlberg chaired, Boyd (1976) used intensive discussion of both hypothetical and real

moral dilemmas in two small college classes. In three months the two classes advanced one-fifth and one-third of a stage, respectively. The first gain was not significant, and the second one was "almost significant" (p. 139). Quite possibly the discussion method does not work. Then again, perhaps the near miss of the second class was partly attributable to a small n (16 and 12 for experimentals and controls, respectively) and to the imprecision of the interview assessment device.

Using a larger n and the more reliable DIT, Panowitsch (1975) conducted a dissertation study comparing an ethics class (enrolling 73) with a logic class (enrolling 28). The first presumably discussed more problems in the moral domain, whereas the second dealt with clearer thinking in general. The first class moved the DIT mean significantly from 41 to 46 P%, whereas the second produced a DIT gain of less than 1 P%, which was not significant.

The purpose of the Panowitsch study was to provide more construct validity for the then freshly developed DIT. The mission was successful. However, the study could not and did not purport to show that classes in ethics raise DIT scores, because students who elected to enroll in ethics rather than logic may have been more predisposed to change their moral reasoning in the first place. Only by randomly assigning students to the two classes could a causal relation be shown.

In order to improve upon both the Boyd and the Panowitsch studies, a larger n, random assignment, and the DIT could be used. This three-part strategy would make it possible to demonstrate that the Kohlberg discussion method,

stressing moral dilemmas, causes more gain in DIT scores than does some other educational strategy. Also, if different discussion leaders having different discussion skills were to be employed, differences in the effectiveness of certain aspects of these skills might be revealed. Relevant here might be an admonition by Rest (1974b):

Unless an intervention is specifically focused on moral problem solving and effecting psychological development, there won't be significant changes in DIT scores over a period of a few months. (page 5-12)

With this admonition in mind I designed the following study, which stresses the major discussion method developed recently by Kohlberg's followers.



## CHAPTER II

### METHOD

Although a single plan for the study had originally been devised, some unforeseen ethical and experimental problems arose, making some rethinking necessary after the study had started. Thus, there was an original plan and a revised plan.

#### The Original Plan

The plan formulated before the beginning of the semester called for a large introductory psychology class to be divided into two sets of discussion groups, each directed by an undergraduate teaching assistant (TA). One set was to use the Kohlberg method for the entire semester. The other set was to use what will be called a question method. The former would presumably raise DIT scores more than the latter.

Unfortunately, however, it became increasingly clear throughout the early part of the semester that the Kohlberg method was not provoking discussion among students. They were reportedly talking mainly to their TAs, if at all, and to each other only minimally. Among the probable reasons for the failure are that students were too unfamiliar with each other, that relative strangers were being asked to discuss among themselves issues of too controversial a nature, and that TAs were as yet too unskilled in the use of discussion techniques. When the two methods are described in detail shortly, you will see that the Kohlberg method requires much more skill from the

TA than does the question method. Ten of the twelve TAs reported that in their Kohlberg groups the morale and satisfaction were noticeably lower than in the groups that used the question method.

Thus, some serious thinking about continuing on the original plan began. The primary purpose of the class was to promote knowledge and critical thinking about psychology in over a hundred students, whereas the secondary purpose was to promote moral reasoning for a doctoral dissertation. To put the second before the first would be to allow the proverbial tail to wag the dog. However, a way around the problem was seen such that indeed more gain in DIT scores might actually result than under the original deteriorating design. Thus, a substantial change was introduced.

#### The Revised Plan

At midsemester, groups originally on the Kohlberg method were switched to the question method, and those originally on the question method were switched to the Kohlberg method. Thus, each student got equal doses of each method, thereby being treated equitably and fairly.

Also, there was new hope for optimism about gain in DIT scores. Under the original plan the hypothesized gain would have been greater under the Kohlberg method, yet it had apparently been introduced prematurely--at a time before student-to-student discussion habits and good rapport had been established. Thus, little DIT gain had probably occurred here. (This conclusion is, of course, conjectural. It would have been

informative to administer the DIT at midsemester as a check, but doing so would have meant administering it three times instead of two. Too much familiarity with the instrument could then have been a problem at the end of the semester.)

Under the original plan the groups that had used the question method had established very effective communication patterns, familiarity, and rapport, due to the nature of the method, to be described shortly. Thus, the question method had primed these groups to the point where they were now ready to talk comfortably among themselves about a variety of deeper, sometimes sensitive, moral issues.

#### Teaching Assistants

The first major logistical step in implementing and designing the study was to select undergraduate TAs one semester in advance of the actual class. Upon starting the class the training program began.

#### Selection

Each psychology teacher at the University was asked to nominate a few students regarded as TA material. Upon completing questionnaire and interview procedures resembling those of Fernald (1973) and upon taking the DIT, these candidates were asked to demonstrate criterion related skills by conducting mock discussions of B. F. Skinner's Walden Two with another applicant and me.

Applicants guided the discussion with given, contrived questions like those later to be written by their own students

under the question method. TA selection emphasized adeptness in using the questions and improvising upon them to keep the discussion going.

From these applicants 12 students were selected as TAs. The mean of their DIT scores was 56 P%, the range running from 45 to 75 P%. (Not included in this range was a score of 28 P% for one TA who permanently took another's place in the early part of the semester. Although extreme in this sense, she did not appear to have extreme effects upon either of her two groups in terms of DIT gain. Neither one was eventually in the upper or lower quartiles of the distribution of 24 group means for DIT gain.)

Of the 12 TAs there were six sophomores, two juniors, and four seniors. The women and men numbered nine and three. They included seven psychology majors and five who were either undecided or were not psychology majors. The mean number of psychology courses taken by them before becoming TAs was 3.6, ranging from 1 to 8. On the University's grading system, where A = 4.0, the mean of the TAs' overall GPAs was 3.4, ranging from 2.9 to 3.8. The mean GPA for all students at the University was in the vicinity of 2.8, about a B-.

#### Training and supervision

Each TA was enrolled in a credit course on the teaching of psychology. The time in class for a TA amounted to two hours of meeting with other TAs and me on the first day of each week and four hours of meeting with his students later in the week. On the first day of the week there were two TA

meetings, the first for the first six TAs, the second for the second six TAs. Thus, TA meetings could be conducted more or less as models of the student meetings to be held later in the week, the same ratio of leader to participants holding in each case.

TAs prepared for each weekly TA meeting by reading the assigned paperback, either Walden Two (Skinner, 1976), Summerhill (Neill, 1960), or The Art of Loving (Fromm, 1956). From this reading the TA then extracted a moral dilemma and appropriate follow-up or probe questions. From the dilemmas and probes submitted by all of the TAs, I selected--simply on a subjective basis--the two dilemmas that appeared to be most promising for discussion at each of the student meetings to be held later in the week.

At each TA meeting we would hold a discussion of either or both of the selected dilemmas and of one or two other dilemmas based on work by Kohlberg's staff. Sometimes I would play the role of the discussion leader and sometimes I would ask one or another of the TAs to be the discussion leader. Overall, TAs became quite adept at this format, often enabling the discussion to continue at length with little or no intervention from me.

Note that I was thus in an advantageous position to assess the extent to which each TA was skillful in the use of Socratic techniques to focus the discussion upon conflict about moral issues. Since TAs knew that I was evaluating this probing ability and all other aspects of their performance subjectively as one component of their course grade, my

observations no doubt had motivational effects. However, as will be seen later, the assessment became useful in answering a research question.

### Experimental Design

Each of the 12 TAs was put in charge of two six-student groups. For each TA the first group used the Kohlberg discussion method for the first half of the semester then switched to the question method, thereby being designated sequence K-Q. The second group reversed the methods, thus using sequence Q-K. Therefore, in a nutshell, there were twelve TAs by two sequences of methods, with six students at each of the 24 intersections or discussion groups.

As a related matter, however, it must be mentioned that one feature from the original design had to be carried over into the revised design, even though no useful purpose was served. Specifically, the original design attempted to maximize the DIT gain under the Kohlberg method by employing several gain producing practices found in the literature. One of them is the maximizing of variation in different stages of moral reasoning in a particular discussion group so that each member can regularly hear moral reasoning one stage above his or her own (except for the member at the highest stage, who would necessarily only hear lower-stage reasoning). Thus, in each of the K-Q groups, variation among the six pretest DIT scores of the students was maximized. That is, from the 72 students who were to be assigned to the 12 groups, the highest 12 students were each assigned to distinct groups, then the

next highest 12 students were assigned in the same manner, and so on. In contrast to this variance maximizing, stratified random assignment in the K-Q groups, the assignment in the Q-K groups was strictly random, so that in some of these groups there was fairly wide variation and in others there was not.

### Procedure

Each week each student was asked to attend a 50-minute lecture and a 50-minute discussion on both Tuesday and Thursday afternoons. Only the discussions were actually required for grading purposes, but three absences were allowed from them throughout the semester without penalty. Appendix B is the course syllabus.

As noted, each student experienced either the K-Q or the Q-K sequence of discussion methods. Early in the semester during a lecture period the two methods were demonstrated in detail by having the TAs play the role of students with me in the role of their TA. Each method will now be described in more detail.

#### Method K(A): Discussion of moral dilemmas

This method, named after Kohlberg and known to students as Method A, is a modification of a method proposed by Galbraith and Jones (1975, 1976) and derived from Kohlberg's theory. Appendix C is the form that instructed students in its use. In essence the method involves presenting students with a moral dilemma that challenges their own level of moral reasoning but that can be more satisfactorily resolved at a

higher level of reasoning. A single dilemma--if broad enough (as, say, the Heinz dilemma, shown already in Table 1)--can challenge any level of moral reasoning (as shown by the 12 Heinz alternatives).

As a brief aside, let me give an example of a typical dilemma and a reaction to it--but from a local citizen instead of a student. Here is the Heinz dilemma, roughly as I stated it on a local, call-in radio show that was dealing with moral education.

If Heinz's beloved wife is dying of cancer, which can be cured by an exorbitantly, unfairly priced drug that Heinz absolutely cannot afford, should he steal it from the druggist if that is the only way to save her?

Upon indicating that most students in discussion say "yes" to show greater value for life than for property, I was attacked by the next caller. He maintained that educators are teaching young people to steal, especially to shoplift, and that to put life above property was "communistic". After he had hung up the moderator remarked some time later that this caller had misinterpreted my remark. The caller then called back to give his name and address and to challenge me to a debate on the radio or in the local newspaper. Rather clearly, some conflict had been created for this man.

Under these circumstances, debating would not, I am fairly sure, have changed his mind or his moral reasoning, but when students can discuss issues like this in a warm, open, non-defensive atmosphere, it would seem possible that moral reasoning could undergo a change. Indeed, in a later survey of opinion about the groups, each student was asked to rate



the extent to which there was a "warm climate, despite disagreements." On a 5-point scale (where 5 = "excellent" and 4 = "good") the mean rating was 4.40, with only one group falling slightly below 4.00.

Under method K a TA would begin each discussion period by taking less than a minute to read aloud short descriptions of three moral dilemmas. All of these issues for the entire semester were contained in a document called the Fireside Fifty, reproduced in Appendix D, and abstracted mostly from Blatt, Colby and Speicher (1974). Upon hearing each dilemma the six students would indicate by a show of hands whether they favored one horn of the dilemma or the other or were undecided. The dilemma that produced the best split in opinions was then used in the discussion a week later, thus giving the TAs a chance to discuss it in the intervening TA meeting. For example, here is the encapsulated description of a dilemma that produced good splits in most of the groups:

If the father in a family with four children becomes paralyzed and the mother still needs sex, should she have an extramarital affair or get a divorce?

After the preliminary voting on the three dilemmas, the remainder of the period was spent on two dilemmas, nearly 25 minutes on each. The first was a fuller, more detailed version of the one selected from the last week, and the second was a dilemma written by one of the TAs, edited by me wherever necessary, and based on the present paperback assignment. An example of such a dilemma is given in Table 2.

## TABLE 2

## EXAMPLE OF A MORAL DILEMMA AND PROBE QUESTIONS

BASED ON WALDEN TWO: FREEDOM VS. CONTROL

Frazier claimed that the most terrifying question that he could ask Castle was, "What would you do if you found yourself in possession of an effective science of behavior" and could then control people? (p. 240) Castle answered that he would dump this knowledge in the ocean so that people would be free.

Frazier then implied that we are all controlled--like it or not--by sources that we cannot identify but mistake for freedom. Thus, Castle would be putting the control into unidentified hands, says Frazier, such as those of the "charlatan, the demagogue," the salesperson, "the bully, the cheat, the educator, the priest."

If total control is a fact of life, should it remain--as it is now--in the hands of the people Frazier mentioned or should it be put in better places?

1. Who should decide what the word "better" means? Can one person ever know what is right for another? How?
2. Is Frazier infringing on anyone's rights by controlling them at Walden? (Didn't they all agree to join under those conditions?)
3. If the world seems to be falling farther and farther away from being civilized, should Frazier give his techniques to the world so that total annihilation can be avoided?
4. Should people strive for freedom if there exists the possibility that there is none?
5. What if 3/4 of the 50 states ratified a constitutional amendment that replaced our present government by a Board of Planners? Would that be just or fair?
6. If control were instituted by constitutional amendment, would anyone ever be justified in breaking any of the laws that controlled him or her? Is breaking laws that a majority support ever justified? Why?

TAs were urged to take the probe questions quite seriously and to memorize them before meeting with their students. Much practice in using probes was gained in TA meetings at the start of each week. The probes were designed to create conflict between moral attitudes or beliefs at the student's own level, to make students question the adequacy of them, and to get students to search for and hopefully find more adequate, consistent, higher-stage moral reasoning. For example, probe 5 in Table 2 seeks to create a conflict between, on the one hand, the conventional belief of Americans in freedom such that each person is free to chart her own destiny and, on the other hand, the conventional belief in law and order as specified by the Constitution. Going further, probe 6 opens the way to a consideration of civil disobedience. If students themselves failed to mention the last concept, then a TA on his toes might introduce it by saying something about the Boston Tea Party or Martin Luther King's letter from a Birmingham jail.

At the conclusion of the discussion of each dilemma the TA was instructed to ask students to summarize the best reasons on each side of the issue. The TA would also sometimes ask if there was a "right" answer, although there was no real attempt to encourage consensus on this point. As part of the overall goal of encouraging more postconventional reasoning, each discussion sought to leave questions in students' minds about which reasons really were the best.

Students were not graded on the quality of their discussion contributions but were graded on a pass-fail basis

on whether they attended, talked a bare minimum, and turned in some written work. Regarding the latter, students were instructed to listen to each other carefully enough to write a brief summary of what any particular student said during the discussion. At the end of each discussion students were assigned the homework task of reporting only one other student's statements, and this student, called the Key Person, was revealed by the TA at the end of the discussion. Upon returning to the next discussion each student was to turn her report in as a Listening Sheet in duplicate. The first copy was graded on pass-fail by the TA, but the second copy went to the Key Person. It was the latter's responsibility to make written responses on each student's Listening Sheet and then return them to the appropriate students as feedback on how well they understood the Key Person. Thus, although not graded, this second Listening Sheet hopefully exerted some peer pressure on students to listen carefully to each other and to present their views clearly.

Method Q(B): Discussion of general issues

This method, named after questions that students themselves wrote for discussion, was known to students as Method B. It evolved from discussions directed by TAs in several of my previous courses. Appendix E is the form that instructed students in its use. In essence the method involves having students carefully formulate and discuss their own questions based on the current paperback assignment.

Before coming to any discussion each student was to read the Summerhill, Walden Two, or Art of Loving assignment. From it he was to select a topic of interest to himself and of likely interest to other group members. On this topic a discussion question was to be written on an index card, so as to encourage conciseness. The question was to integrate the paperback with the course text (or other readings) or with real life; preferably all three were to be integrated. Questions were to reflect thinking in their formulation, were to be as controversial as possible, and were to have no clearcut easy answers. Students were given the following question as an example that meets the specified criteria:

Frazier's sheep used to avoid the fence because it was electrified. However, even after he permanently "unplugged" it, they and even their offspring avoided it. Is this "tradition" in any way like any human traditions, such as sanctions against premarital sex or adultery? Why or why not?

For examples of questions that students actually wrote, see Appendix F.

Having written the question, the student would then go to the discussion meeting. Here he would take his turn in a circle, monitored by the TA. A turn consisted of having the student read off the question to colleagues and then--with the help of the TA when necessary--keep the discussion going for about 8 to 10 minutes. Thus, each of the six students would get one turn in the 50-minute discussion period.

Under this method the grading was essentially the same as under the other method. That is, discussion contributions were not graded, but attendance, talking at least

minimally, and turning in a question card were graded on a pass-fail basis. It was thought that peer pressure would in most cases insure that publicly read question cards would be of sufficient quality so as not to be a waste of an entire discussion group's time.

### Subjects

In the introductory psychology course that supplied the Ss there were 98 freshmen, 31 sophomores, 11 juniors, 3 seniors, and 1 graduate student, totaling 144. There were 94 women and 50 men. For 98 students the course was elective, but for 46 students it was not. Since the class and group meeting times were late in the afternoons--an unpopular time slot--most students had been assigned to this particular section of the course, rather than selecting it. There were 25 psychology majors. Although specific SAT scores were not readily available, the students came from a student body in which the verbal mean was 477 and the quantitative mean was 539, both means above the national means.

## CHAPTER III

### RESULTS

This chapter will describe the measures recorded. Then it will present the summary values actually obtained on these measures, and it will show some of the relations among them.

#### Measures Recorded

Cast in research terms, the study had two purposes. The primary one was to demonstrate a possible causal relation between teaching methods and resulting changes within students. The secondary one was to explore other variables that might be useful in explaining any primary relations that materialized; hopefully, suggestions for further research would also emerge.

#### Primary variables: Mostly experimental

Two classes of variables--the independent and the dependent ones--were experimental. In addition, a covariate was included.

Independent variables--The original plan had called for the two methods themselves--K and Q--to be the two values on a qualitative independent variable. However, the revised plan made the two sequences of methods--K-Q and Q-K--the two values on a methods sequencing independent variable.

In addition, TAs themselves were twelve values on another qualitative independent variable. Yet if TAs could be ordered on some metric basis, it would define a quantitative

independent variable, thus making trend analyses possible. Two such quantitative variables on TAs did indeed suggest themselves on a theoretical basis. The first was the preliminary DIT scores of the TAs, as reported earlier. The second was the ability of TAs to ask their students probing follow-up questions that focused on moral issues.

This skill, henceforth called probing ability, was originally to have been rated by a master TA who would have visited all TAs in their respective student discussion groups. A special master TA did in fact perform this duty, but only sporadically, because she also was needed to take the place of TAs who were absent or--in one case--quit school. Thus, rather than rely upon incomplete ratings of probing ability, I myself rank ordered the TAs on this ability. My ratings were based on what I could observe in TA meetings, where each TA had been given a number of opportunities to play the role of a discussion leader. Although it is possible that a TA's degree of probing ability under the gaze of her peers and teacher might not transfer to the actual student discussion groups, the ratings that I made of this ability were later to suggest a rather revealing relation.

Dependent variables--During the first class meeting students completed the DIT, described previously. Then three months later at the end of the semester, this scale was administered again. Thus, the amount of change in moral reasoning as measured by the DIT, could be assessed.

The remaining category of dependent variables was ratings of satisfaction by students. First, they were asked



to give an early, tentative rating on a 5-point, Likert scale of the overall educational value of the first method that they were using--either K or Q. Thus, the simple main effects of methods K and Q when each was administered first in the sequence were assessed. Then at the end of the course each student rated the overall educational value of either the K-Q or the Q-K sequence; thus, the main effects of the two sequences were assessed. Finally, at the end of the semester on this same Likert scale each student rated methods K and Q separately. All of the Likert ratings were understood by students to be an assignment that fulfilled part of their laboratory participation requirement. Students put only their Social Security numbers, not their names, on the rating forms, which were guaranteed to be kept confidential.

A covariate: Dogmatism--A student's stage of moral reasoning, conceived in Kohlberg's terms, could be considered to be a specific complex characteristic of the student's entire mental system of beliefs and disbeliefs about everything--philosophy, ethics, religion, politics, science, practical matters, and so on. Such a system can be portrayed as falling on a continuum from open to closed, according to Rokeach (1960). A system is open to the extent that, first, relations both between and within all parts of the entire belief-disbelief system can be seen for what they are and, second, there is not an absolute, irrevocable rejection of disbelief subsystems--say, fascism, atheism, or whatever.

From these two related characteristics of the open mind it would seem to follow that more openness would imply a

greater likelihood of seeing true conflicts or contradictions within the belief system, which would include the beliefs characteristic of one's moral stage, such as the stage-4 belief that authorities should always be obeyed. It would also seem to follow that the discovery of conflicts within one's own belief system would, in the open mind, make serious examination and possible acceptance of previously rejected disbelief subsystems more likely. If these disbelief subsystems include both lower and higher stages and if, as Kohlberg asserts, the higher stages are less conflict laden, then the more open the mind, the more likely it is that discussion-aroused conflicts will induce movement to higher stages, that is, the more likely a gain in DIT scores, operationally speaking.

Rokeach measures openmindedness with the Dogmatism Scale. If, as Kohlberg maintains, the higher the stage, the more cognitive-moral development, that is, the more movement, then dogmatism as the opposite of openmindedness should correlate negatively with postconventional thinking or DIT scores. Such a relation was indeed obtained in a pilot study that I performed on a previous introductory psychology class, which yielded a Pearson correlation of  $-0.48$  ( $p < .05$ ) between the Dogmatism Scale and P% of the DIT.

It would thus seem that any cognitive movement within a particular semester should also correlate negatively with dogmatism. It would also seem likely that by removing the dogmatism component from DIT scores or DIT gain scores, as in an analysis of covariance, the power of the statistical test should be improved. For these two reasons the 40-item

Dogmatism Scale (Form E; Rokeach, 1960) was administered to each of the 144 students during the first class meeting.

Secondary variables: Mostly exploratory

In order to examine some other possible relations that could be either explanatory or heuristic, the research effort gathered various ratings--by students, by TAs, and by me.

Ratings of discussions by students--Students rated different characteristics of the TAs and the two discussion methods. All such items are shown in Table 3.

Ratings by TAs--The TAs were also asked to rate students on two aspects of their discussion performance. Having filled out the DIT and having scored other unidentified DITs, each TA was asked to rank order the six students in one of her groups based on what she estimated the DIT scores to be. TAs never knew the actual scores and thus had mainly--perhaps only--the students' statements in discussions as a basis on which to judge. TAs did this only for groups that had method K first, that is, the groups in which the DIT dispersion had been maximized. The ratings were made the week after method K ended for these groups.

The second estimate of students by TAs was made after each meeting. Specifically, students' Listening Sheets (see Appendix C), although given pass-fail marks for grading purposes, were also rank ordered on the basis of the TA's perception of the students' accuracy in comprehending what the Key Person (see Appendix C) had said. These rankings were then averaged over all of a student's Listening Sheets to give an overall picture of how well he understood the moral reasoning of others.

TABLE 3

STUDENT EVALUATION OF INSTRUCTION:  
THE DISCUSSION GROUPS

For purposes of grading your TA fairly and conducting valid research it is most important that you answer completely and accurately. As your instructor I give you my word that this sheet will be held in strictest confidence. Also, your name will never appear on it.

Please rate each of the following items in terms of general quality by circling the appropriate number:

<u>Your TA</u>	<u>Excellent</u>	<u>Good</u>	<u>Average</u>	<u>Poor</u>	<u>Awful</u>
1. Knowledge of psychology	5	4	3	2	1
2. Friendliness or rapport	5	4	3	2	1
3. Skill in promoting discussion	5	4	3	2	1
4. Overall effectiveness as a TA	5	4	3	2	1
<u>Method A (Talk, then listening sheets)</u>					
1. Relevance to the lives of students	5	4	3	2	1
2. Exchange of ideas between students	5	4	3	2	1
3. Ability to cause disagreements (whether friendly ones or not)	5	4	3	2	1
4. Warm climate, despite disagreements	5	4	3	2	1
5. Promotion of new opinions in you	5	4	3	2	1
6. Fostering of personal growth	5	4	3	2	1
7. Value of the listening sheets	5	4	3	2	1
8. Overall educational value	5	4	3	2	1
<u>Method B (Question cards, then talk)</u>					
1. Relevance to the lives of students	5	4	3	2	1
2. Exchange of ideas between students	5	4	3	2	1
3. Ability to cause disagreements (whether friendly ones or not)	5	4	3	2	1
4. Warm climate, despite disagreements	5	4	3	2	1
5. Promotion of new opinions in you	5	4	3	2	1
6. Fostering of personal growth	5	4	3	2	1
7. Value of the question cards	5	4	3	2	1
8. Overall educational value	5	4	3	2	1

Rating by course instructor: Essay question on loving--Since principled moral reasoning might seem to be related to Erich Fromm's concept of loving, as expressed in The Art of Loving, perhaps the relation could be operationalized. All that was needed was a measure of students' understanding of Fromm's concept. This was obtained from students by scoring their responses to the following essay question from one of their exams:

What similarities (if any) and differences (if any) do you see between what Fromm means by loving and what he sees as the average person's meaning of love?

My student assistant, having exceptionally high ability, and I both read Fromm to come up independently with the main points for a good answer. Our interrater reliability in grading a random sample of 12 essays from the 144 was .91, within a confidence interval from .67 to .98 ( $p < .05$ ). Our main points were three. First, in loving there is only giving without expecting something in return, whereas in love there is an expectation of mutual exchange. Second, loving is an active, voluntary process, whereas love is something passive that you involuntarily feel for someone who is attractive. Third, loving extends to everyone, whether or not in one's own family, group, or nation, whereas love is more or less in direct proportion to your actual social contact with or knowledge of another person.

Each of the three main concepts on each student's paper was awarded percentage points up to some maximum, which was 35 for the first, 35 for the second, and 30 for the last.

Partially articulated answers, indicating partial understanding, were subjectively awarded partial credit.

### Values and Relations Found

SPSS, the computer routine by Nie and colleagues (1975), calculated most of the following statistics, consisting of summary values, measures of relation, and tests of significance. The analysis included both experimental and exploratory variables. (The N is nearly always slightly less than 144 due to three students who dropped the course, another whose deafness severely hampered communication, and a very few others who could not be later tracked down to remedy unusable or missing data.)

### Experimental findings

The results that address themselves to identifying causal relations can be considered under the headings of DIT gains, dogmatism, and general ratings by students.

DIT gains--For the K-Q sequence of methods the DIT yielded pretest and posttest means of 43.03 and 44.20 P%, respectively. For the Q-K sequence the respective means were 42.27 and 45.53. Thus, for the K-Q and Q-K sequences the mean DIT gains were 1.17 and 3.26 P%, respectively. However, this difference between the two sequences in terms of gain is not significant,  $F = 1.48 (1,136)$ ,  $p > .05$ .

The two prevalent quantifiable aspects of TAs, namely, their DIT scores and their probing ability ranks, were related to DIT gains as follows. First, each TA's DIT score was paired

with the mean DIT gain for his or her K-Q group, then across all TAs these twelve pairs were correlated to yield a nonsignificant correlation of  $-.09$ . The same analysis was performed on the Q-K groups to yield another nonsignificant correlation of  $.06$ .

Second, each TA's probing skill rank was paired with the mean DIT gain for his K-Q group, then a Pearson product-moment correlation (the only type used in this study) was computed between the two measures but was not significant. When I repeated this procedure for the Q-K groups, however, significance appeared in the form of a correlation of  $.65$ ,  $p < .05$ . (Accordingly, the correlation between the Q-K students' DIT gains individually and their respective TAs probing skills was  $.23$ ,  $p < .05$ .) This result must be regarded cautiously, because the level of significance is not strong, and the analysis was somewhat post hoc, thereby making Type I errors more likely.

Dogmatism, the DIT, and covariance analysis--The correlation between the pretest and posttest of the DIT was  $.66$  ( $p < .01$ ), but that between the Dogmatism Scale and the DIT posttest was  $-.27$  ( $p < .01$ ). Since both the pretest and the Dogmatism Scale correlated significantly with the posttest, either one of the two covariates could be used as a covariate in an analysis of covariance on the posttest. In neither of the two analyses of covariance that were then performed, however, did significance obtain: for the pretest as the covariate on the posttest,  $F(1,135) = 1.45$ , and for dogmatism as the covariate on the posttest,  $F(1, 135) = 1.78$ . Thus, by covariance analysis the two sequences of methods did not differ from

each other. (The use of these two covariates was suggested by Harris, 1975, pp. 22-23.) Finally, dogmatism was used as a covariate on still another dependent variable, DIT gain scores, but here again there was no significant difference between sequences in the change they produced,  $F(1,135) = 2.13$ .

Student ratings of discussion: General--In the preliminary rating of discussions the students using method K and those using method Q gave the two methods mean ratings of 3.98 and 4.22, respectively (where 3, 4, and 5 were defined as "average," "good," and "excellent"). This difference is significant,  $F(1, 110) = 4.40, p < .05$ .

As indicated earlier, each student at the end of the course rated either the K-Q or the Q-K sequence (in addition to rating both methods K and Q separately as shown in Table 3). In this final rating of the two sequences--that is, what for each student was an overall rating of all discussions throughout the semester--the K-Q students gave their meetings a mean assessment of 4.16 (a bit more than "good"). In contrast, the Q-K students awarded their groups a mean of 4.58 (more "excellent" than "good"). These means are significantly different,  $F(1, 137) = 9.87, p < .01$ .

Looking more specifically at the final assessment of each method, we see that method K was rated by the K-Q students and the Q-K students at means of 3.67 and 3.97 (both mostly "good"), respectively,  $F(1, 135) = 4.92, p < .05$ . The direction of this difference was the same for the mean ratings of method Q, 3.42 (above "average") and 3.99 (a hair from "good"),  $F(1, 135) = 14.07, p < .001$ .



### Exploratory findings

To explore various leads the analysis computed correlations between various ratings made by students, by TAs, and by me.

#### Student ratings of discussions and TAs: Specifics--

For each discussion group the mean of each of the twenty specific items in Table 3, showing different aspects of the TA and the two discussion methods, was computed. For all of the twelve Q-K discussion groups these means and the mean DIT gain were submitted to a correlation analysis, yielding a correlation matrix, whose entries showed the correlation between each of the 210 possible pairs of variables. Thus, for the more promising Q-K sequence it was possible to look for correlates of gains in the DIT.

In this matrix three preliminary findings are of note. First, DIT gain correlated significantly with none of the twenty items. Second, nearly half of the eight items on method Q correlated significantly with each other, but this is probably at least partly artifactual due to the halo effect of trying to remember the specifics of a method that each student had last used more than six weeks previously. Third, the only significant correlation between any of the method K items and any of the method Q items was that between the ability of each method to cause disagreements:  $r = .72, p < .01$ .

These three types of correlations involved items that were then removed from the correlation matrix in order to obtain a more parsimonious, understandable, and useful table. The reduced matrix is shown in Table 4.

TABLE 4

CORRELATIONS BETWEEN SELECTED ITEMS RATED BY  
STUDENTS WHO USED THE Q-K SEQUENCE OF METHODS

	Teaching Assistant			Method K		
	Friendliness or Rapport	Promotion of Discussion	Overall Effectiveness	Relevance to Life	Exchange of Ideas	Listening Sheets
TA: Friendliness or Rapport		86*	86*			
TA: Promotion of Discussion	86*		86*		68	
TA: Overall Effectiveness	86*	86*			69	
MK: Relevance to Life						72
MK: Exchange of Ideas		68	69			
MK: Listening Sheets				72		

Note--Decimal points are omitted. Correlations for which  $p < .001$  have asterisks (\*), whereas those for which  $p < .01$  do not. Nonsignificant correlations are not shown.

TA ratings of students--For each TA a correlation coefficient was computed between (a) his estimated ranks of his six method K students on the DIT and (b) those students' actual DIT scores. The weighted mean of the twelve correlations--obtained by an r-to-z transformation (McNemar, 1962, p. 140)--was 0.35, which was not significant, and they ranged from 0.93 to -0.81. The twelve correlations were then correlated with the TAs' DIT scores. The resulting correlation was -0.12, which also was not significant.

A correlation was computed for the mean rank of each student's entire set of Listening Sheets (throughout the entire K half of the semester) and the amount of DIT gain for each student, both across all of the Q-K students. (I explored the Q-K relations rather than the K-Q ones because the former seemed more promising in revealing what was shown to be more DIT gain. Also, the Q-K N of 72 provides very stable estimates, almost as stable as an N of 144, and probably sufficient for heuristic purposes.) This correlation between the Listening Sheets and the DIT gain was not significant. On the other hand, this mean rank of each student's collected Listening Sheets was found to correlate with the multiple choice scores on the course exams and with the essay scores on these same exams. These correlations were 0.40,  $p < .001$ , and 0.28,  $p < .01$ , respectively.

Instructor rating of essay question: Postconventional loving?--On the essay question that contrasted love with loving, the mean score was 55 percent, the range running from 8 to 100 percent. These scores did not correlate significantly with

either the pretest or the posttest of the DIT. However, the scores did correlate with the rankings on the Listening Sheets, this correlation being 0.27,  $p < .01$ .

## CHAPTER IV

### DISCUSSION

It might help to first briefly pull some earlier strands together, then unravel them for an extended, closer look. In this fourth and last chapter the main hypotheses and speculations of the first two chapters will be summarized in relation to the findings of the third. Then a longer discussion will offer comments and speculations.

#### Main Hypotheses and Results: A Rundown

Originally the major hypothesis was that an entire semester of the Kohlberg method would produce more gain in moral reasoning--DIT scores--than would the question method, a previously proven educational tool. Early ratings of both methods by students, however, showed significantly less satisfaction with method K; other signs were also unfavorable to its effectiveness.

Upon revision the major hypothesis then became that DIT scores are raised more over the entire semester by the Q-K sequence of methods than by the K-Q sequence, method Q being an effective warm-up for method K. This revised hypothesis was not confirmed directly. Some indirect, tentative support did appear, however, for the superiority of the Q-K sequence, because it did act as an effective mediator for a theoretically and empirically expected relation between Socratic skill of the TA and DIT gain in her students. The relation was weak but significant.

An adjunct hypothesis was that dogmatism hampers moral growth; operationally speaking, the lower the dogmatism score, the higher the DIT gain. This hypothesis was not confirmed.

#### Conditional Changes in DIT Scores

Previous intervention studies that have tried to produce gains in the DIT have met with mixed success. A review of fourteen such studies (Lawrence, 1977) found that about half of them produced significant DIT gains. The present study may be typical of all fourteen as a group, because while direct, convincing effects of the intervention were not shown, indirect effects were.

That is, under the appropriate conditions, gains occur. First, starting out cold with method K did not seem to produce either student satisfaction or much DIT gain, but preceding it with an ice-breaker like method Q produced an average gain of about three-and-a-quarter percentage points of principled reasoning on the DIT. This showing may not really fall too far short of the findings from other studies of similar duration, because they show gains of similar magnitude--usually about four percentage points. It must be remembered, however, that the three-and-a-quarter point gain for sequence Q-K was not significantly better than the one-and-a-sixth point gain for sequence K-Q.

One reason that the DIT gain under sequence Q-K was not actually more impressive may have been that one of the discussion vehicles, The Art of Loving, turned out to be rather inappropriate. In final student ratings of the educational

value of course materials, where 5 = "excellent," 4 = "good," and 3 = "average," Summerhill and Walden Two got means of 4.34 and 3.78, but The Art of Loving received a mean of only 3.23. Possible reasons for its inappropriateness were suggested from time to time by students and TAs and by their written, open-ended evaluations of the course. For example, some students thought it sexist that in Fromm's characterization of different types of love he felt that motherly love and fatherly love were, respectively, like what is in essence Roger's unconditional and conditional positive regard. Moreover, the book is probably too theoretical and not tied down enough to everyday life, as are the other two books. Therefore, Fromm's book may have been too abstract for these students to relate to and to understand, and thus it may not have generated stimulating discussions.

Another reason that DIT gains may have been smaller than they could have been was the nature of the dilemmas themselves. Recall that about half of them were adapted from a set produced by Kohlberg's group and the other half were adapted from the paperbacks by the TAs under my editorial supervision. Unknown to me until late in the course, some of the Kohlberg group's dilemmas had already been discussed by some students in high school. Also, comments were sometimes heard that the dilemmas were not "relevant" enough to the lives of students. This criticism is, of course, a subjective one assessed by each student on his own and by each reader upon turning to Appendix D and sampling some of the dilemmas. In contrast, the dilemmas on the paperbacks were probably seen as a bit

more relevant, except, of course, for some of the dilemmas from The Art of Loving. Here again, these later dilemmas may have held down the DIT gain from being all that it could have been.

As suggested earlier, the warm-up effect of the question method upon the Kohlberg method may be the major reason for the possible superiority of the Q-K sequence over the K-Q sequence. Another possible contributing factor, however, would be improved skill of the TAs in using the Kohlberg method as the semester progressed. Thus for the K-Q groups method K would have been used in the hands of the relatively unskilled, whereas for the Q-K groups the opposite would have been the case, thereby giving an added boost to DIT scores under this later sequence.

Although this explanation involving TA skills is only hypothetical, it is indirectly supported by what is probably the major finding of this dissertation. Namely, provided a sufficient warm-up method precedes the Kohlberg method, the more Socratic skill a TA exhibits the more her students will gain in DIT scores. The gain in this study under the Q-K sequence ranged from -2.50 P% for the TA whose group showed the least gain through 3.26 P% as the mean for all groups to 9.33 P% for the TA whose group showed the most.

This main result must be interpreted cautiously, however. As mentioned before, this analysis was post hoc and only turned up a barely admissable level of significance (.05). In addition, a caution of another sort must be noted. Although it would seem likely that a TA's Socratic skill does lead to



DIT gains, that inference is not really warranted by the data. All that has been shown is that TAs probably produce significantly different amounts of DIT gain and that these gains are linearly related to the TAs' Socratic skills. It could be, however, that there is some other TA characteristic--such as intelligence--which is confounded with Socratic skill and is the real cause of the DIT gains. Or perhaps there is some constellation of causes.

It must also be admitted that there is still another possible direction in the causal chain. Perhaps certain groups of students "clicked" better, that is, the mix of personalities produced more productive discussions. If this activity encouraged the TA to then play more of an active or involved role and led him to put more preparation into the class and to develop greater skill, then his Socratic skill might have been a growing byproduct of the characteristics of his group. After all, my ratings of the TAs was cumulative over the entire semester. This group-caused explanation, however, seems less parsimonious and thus less likely than the TA-caused interpretation that I prefer.

The Socratic-gain finding corroborates one result of another study. Colby, Kohlberg, and associates (1977) examined several key moral discussion variables in about twenty Boston high school classrooms. Teachers had been trained in the use of the same moral discussion method employed in this dissertation. During the academic year 24 discussions, each employing a single dilemma, occurred in each class, where meetings were periodically observed by the researchers. Interviews of each

student before and after this program assessed gain in moral reasoning on Kohlberg's moral maturity scale--not the DIT. A criterion of gain was then used to classify each entire class as having gained or not. In all but one of the gain classes and in only one of the no-gain classes teachers had been judged by the researchers to use probe questions effectively. That is, these exemplary probers could improvise with their own probes in addition to those of the lesson plan, and such teachers could focus on moral issues and reasoning rather than non-moral issues and choices. Thus, both the Colby study and the present one seem to point to the importance of the discussion leader's probing skills in promoting gains in moral reasoning--whether measured by Kohlberg's subjective method, which has been criticized by Kurtines and Greif, or by Rest's hopefully improved objective method, the DIT.

Although it would seem that a TA's DIT score would be related to her probing skills and thus to the DIT gain for her group, the present study does not confirm this notion, since no relations were found involving the DIT scores of the TAs. In fact it is a bit incongruous to note that in the less effective K-Q sequence of methods there resulted in one group a DIT gain of 11.33 P%, the highest mean gain for any group under either sequence, but the TA's DIT score was 45 P%--the lowest of all of the regular TAs! My own observations of her discussions with other TAs and with her students showed that she thrived in stirring up controversy. Interestingly enough, at the beginning of the semester her score of 45 P% was just one point above the mean of her group's, 44 P%. Thus, in

going up to 55 P% during the semester, her students were apparently not merely emulating her but were grappling with issues among themselves through her catalytic assistance. It would also seem likely that her score advanced substantially, too, but a posttest of TAs was not administered because by then they were familiar with the scoring of the DIT, thereby removing themselves from the category of legitimate testees.

One final point about the relation of teachers' level of moral reasoning and that of their students is in order. After administering the DIT to graduate students who were education majors and after finding that their scores were significantly lower than those of students in other fields, Bloom (1976) asked whether teachers relatively low in post-conventional reasoning can promote it among their students. The present study suggests that the answer may be "yes," at least in some cases.

#### Dogmatism and DIT Scores

Confirmed was the significant negative relation between dogmatism and DIT scores, as expected from the pilot data reported earlier. Not confirmed, however, was the hypothesized negative relation between dogmatism and DIT gain.

A possible explanation for the latter finding has been offered in another context by Kemp (1957). In his dissertation, chaired by Rokeach, no relation was found between dogmatism and change on the Allport-Vernon Scale of Values (1931), which taps attitudes that are theoretical, economic, aesthetic, social, political, or religious. Kemp explains that values--

including moral ones, by implication--may either change in both closed and open minds or may not change in both types. More specifically, the closed mind may change a value due to a change in conformity to authority, and an open mind can change a value on a more "genuine" basis, a deeper understanding. Conversely, a closed mind can stick to a given value through rigidity, and an open mind can retain a value through intellectual, not dogmatic, conviction--as in the case where no appropriate or relevant reasons for change have been offered.

Thus, the closed-open, changed-unchanged explanation could describe the four quadrants of a dogmatism-by-DIT-gain scatterplot. The filling of all four quadrants constitutes a nonsignificant correlation, as was indeed found in the present study.

#### Student Satisfaction with Methods

In the early part of the course, when students had been exposed to only one method, those groups on method Q liked it significantly better than their counterparts liked method K. In fact, the majority of students on Q had rated it as "excellent." This result is the same as the previous year's assessment of an entire introductory psychology course that used essentially the same device--an earlier form of method Q--in small discussion groups led by TAs; most students indicated in open ended responses that the groups were what "made" the course. The fact that in both years students wrote their own questions under method Q probably helped to sustain active student involvement and satisfaction. Perhaps for this

reason and others already discussed, it is not too surprising that the students on the alternative plan--method K--rated it less favorably at about "good."

Still another later measure of student satisfaction can be used to affirm the suggested but nonsignificant superiority of sequence Q-K and its role as a mediator in the significant linear relation between Socratic skill and DIT gains. Students on the Q-K sequence rated their discussions overall--that is, both methods--as closer to "excellent" than good, but the K-Q students rated discussions overall as roughly "good;" recall that this difference, reported earlier, was significant. Thus, the same pattern of results from the last paragraph appears again in this one, thereby suggesting that once the die is cast in the early part of the semester, little can be done to change student perceptions of the educational value of a particular experience, such as a course, even when the educational methods are changed. This notion is supported by experienced instructors, who stress the pivotal importance of making a good first impression upon students. In fact, in a slightly removed yet still related field, that of personal attraction in social psychology, the literature is replete with evidence for the importance of first impressions. As one contemporary slogan from the everyday world asserts, "You never get a second chance to make a first impression."

An alternative explanation of the different results--in terms of both student satisfaction and DIT scores--between the two sequences could be that the difference was caused by the deliberate maximizing of differences among DIT pretest

scores in each of the K-Q groups but in none of the Q-K groups. Yet this strategy should have made K-Q more effective than Q-K, but that did not happen. Evidence for this claim is suggested by the same study cited earlier, that of Colby, Kohlberg, and associates (1977). They found in their study of the Boston classrooms that the greater the variation in pretest moral reasoning within a classroom, the more likely it is that growth in moral reasoning will result. However, the result from this part of their study approached but did not reach significance.

#### Exploratory Findings and Implications

Some of the many ratings that were gathered did suggest some research paths to greener pastures. Let us take a look.

Student ratings--Somewhat unexpected was the fact that none of the student rated aspects of either their TAs or their classroom group discussions were significantly related to gains on the DIT. It is possible, of course, that students were not sufficiently sensitive to these aspects to assess them validly, but one would think that for an aspect such as the warmth of the discussion climate, students would have some feel for it, and it would seem that this aspect is important (Wilson, 1970).

However, students did seem to be sensitive to the warmth or friendliness of their TA. This quality, as shown by the first quadrant of Table 4, clusters with the TA's rated promotion of discussion and his or her rated overall effectiveness. Thus, all three qualities are essentially the same one

and will hereafter be referred to as the TA's overall effectiveness.

This quality, overall effectiveness, was strangely circuitously related to DIT gain, however, and played an important heuristic role in leading me to the significant relation between Socratic skill and DIT gains in the Q-K students. Here is the key. In groups that rated their TAs low in effectiveness the DIT gains were always low, but in groups that rated their TAs high in effectiveness, the DIT gains could be either high or low. Two things occurred to me.

First, groups who rated their TAs high were perhaps saying that TAs promoted discussion, regardless of whether the discussion focused on moral issues or not. For example, if the Heinz dilemma (see Table 1) had ever been discussed, the verbal exchanges could be quite lively regardless of whether or not morals were discussed. The TA can concentrate on the moral issue of whether stealing is ever right or on the nonmoral matter of clever ways by which Heinz could get the money, such as going to a news reporter to make a public appeal for funds. The TA who stays on moral issues is defined for Kohlbergian purposes to be probing more effectively than the TA who does not. The TA who probes effectively about moral issues creates the moral conflicts said to be necessary for moral growth, whereas the TA who does not probe effectively does not promote this growth.

Second, groups who rated their TAs low probably meant that their TA could neither probe nor keep the discussion

going on nonmoral matters. Thus, none of these latter TAs stimulated moral growth.

From these two speculations it then occurred to me to look at how my ratings of probing ability related to DIT gain. Sure enough, the expected relation did obtain, but only for the Q-K sequence.

Table 4 also shows that TA effectiveness is fairly well related to the exchange of ideas under the TA. This condition would seem to be necessary but not sufficient for the growth of moral reasoning.

Finally, Table 4 shows that students who felt that method K was relevant to their lives tended to take the Listening Sheets fairly seriously. Conversely, students for whom method K did not seem very relevant saw less value in the Listening Sheets.

TA ratings of students--A most curious finding for which I have no explanation is the fact that TAs were not able to rank order their K-Q students in terms of estimated DIT scores. Yet these rankings were requested only after each TA had listened to each student discuss dilemmas for a half semester (18 dilemmas in all, whereas the DIT consists of only 6 dilemmas). Also, the six students for each TA had been assigned in such a way as to maximize the variance between them on DIT scores, and this arrangement should have made them more discriminably different.

Moreover, it would also seem that the more postconventional a TA's thinking--as reflected in a higher DIT score--the more accurately she could recognize the extent to which



each of her students uses postconventional reasoning. Yet this hypothesis was not confirmed either.

In short, substantiation of both of these hypotheses could have offered the DIT increased support in terms of its construct validity. Yet they did not!

The essay question on loving: A key to the postconventional?--Another surprise--to me, at least--was the empirical failure to confirm a relation between postconventional thinking and comprehension of Fromm's concept of loving. A possible explanation for this lack of correspondence is that as global concepts the two do not correspond. Yet there may be certain elements within each global concept that do correspond between the two, but the harmony of isolated pairs may be drowned out by noise from too many other cacophonous pairs.

For example, take a student who correctly comprehends one of three main aspects of Fromm's loving by writing that it "extends to everyone, whether or not in one's own family, group, or nation." It seems likely that this student in taking the DIT and reading the Heinz dilemma would also be likely to check off the postconventional statement, "Would stealing in such a case bring about more total good for the whole society or not? (emphasis added)"

Perhaps in future work a content analysis of each student's written answer to the essay question could be submitted to a multiple regression analysis in which the 21 postconventional statements on the DIT are the predictors. Then those aspects common to Kohlberg's global postconventional concept and Fromm's global loving concept could be identified

empirically. However, such a project would seem outside of the scope of this dissertation and should thus perhaps be pursued elsewhere.

#### Some Final Words: Early Steps

It was Lao-tsu, the ancient patron saint of Taoism, who is supposed to have first observed, "A journey of a thousand miles must begin with a single step." Such a maxim seems particularly appropriate as a terse characterization of this dissertation for three rather unrelated reasons.

First, before taking an extended journey of many weeks with the Kohlberg method, interstudent communication habits and rapport should be established as an early step. The question method seems ideal for this ground breaking purpose, whereas the Kohlberg method does not. That is, the question method serves to prime the pump to get ideas about moral issues flowing freely and openly.

Second, the dissertation itself is only a first step on what could be a life-long journey through a research program in moral education. The first step is less bold than originally intended, however, because the present study could more aptly be described as a pilot study for a bold first step. Instead of clearly showing causal relations of major importance, this study seems more useful in clarifying the conditions under which the causal relations are likely to materialize. More specifically, the Kohlberg discussion method would seem to produce DIT gains when preceded by an appropriate warm-up strategy such as the question method, but it would appear

that TAs should have adequate Socratic skills in order to lead the discussions productively.

Third, a final reason for the appropriateness of the Taoist maxim to this dissertation is that the longest and most profound journey facing the civilized inhabitants of any planet--particularly a small one supporting about four billion persons--is the devising of a cooperative code by which all dwellers can live fairly and justly. Whether we call this task the advocacy of justice, the promotion of morality, or whatever, the state of our contemporary world, which is fairly well described as an international anarchy, would strongly suggest that an overwhelmingly large number of steps toward justice must still be taken. The present moral education movement in our country is only one of the first of these steps. The "Kohlberg bandwagon," as it has been called (Fraenkel, 1976), is probably only a precursor, yet it seeks to promote moral reasoning that will minimize conflicts among a maximum number of people.

One of the bedrocks upon which Kohlberg lays his post-conventional concepts is Kant's categorical imperative. A universal, unconditional maxim, it asserts that the only moral principles upon which to justify actions are those which we would be willing to see universalized. For example, by universalizing patriotism in war, we must say that we actually want the enemy to try to destroy us; but by universalizing a contrasting concept, pacifism, we are less contradictory. Still, although moral philosophers generally agree that Kant's position is necessary for a viable morality, there is less

agreement that his thinking is sufficient, for various reasons beyond the scope of this discussion (Frankena, 1973). Yet it would seem to be a sound first step down the path of moral education.

## REFERENCES

- Allport, G. W., and Vernon, P. E. A study of values. Boston: Houghton Mifflin, 1931.
- Blatt, M. M., Colby, Ann, and Speicher, Betsy. Hypothetical dilemmas for use in classroom moral discussions. Cambridge, Mass.: Harvard University, 1974.
- Blatt, M. M., and Kohlberg, L. The effects of classroom moral discussion upon childrens' level of moral judgment. Journal of Moral Education, 1975, 4, 129-161.
- Bloom, B. S. (Ed.) Taxonomy of educational objectives. Handbook I: Cognitive domain. New York: McKay, 1956.
- Bloom, R. B. Morally speaking, who are today's teachers? Phi Delta Kappan, 1976, 57, 624-625.
- Boyd, D. R. Education toward principled moral judgment: An analysis of an experimental course in undergraduate moral education applying Lawrence Kohlberg's theory of moral development. Unpublished doctoral dissertation, Harvard University, 1976.
- Brown, Roger, and Herrnstein, R. J. Psychology. Boston: Little, Brown and Company, 1975.
- Calder, B. J., and Ross, M. Attitudes and behavior. Morristown, N.J.: General Learning Press, 1973.
- Campbell, D. T., and Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin, 1959, 56, 81-105.
- Colby, Anne, Kohlberg, L., Fenton, E., Speicher-Dubin, B., and Lieberman, M. Secondary school moral discussion programmes led by social studies teachers. Journal of Moral Education, 1977, 6, 90-111.
- Cooper, D. The analysis of an objective measure of moral development. Unpublished doctoral dissertation, University of Minnesota, 1972.
- Dewey, John. Democracy and education: An introduction to the philosophy of education. New York: Free Press, 1916.
- Ennis, R. H., and Millman, J. Cornell Critical Thinking Test. Urbana, Ill.: University of Illinois, 1971.

- Fernald, P. S. The selection and training of undergraduate teaching assistants for large personalized instruction courses. Newsletter of the Division on the Teaching of Psychology, American Psychological Association, March, 1973, 3-5.
- Fraenkel, J. R. The Kohlberg bandwagon: Some reservations. Social Education, 1976, 40, 216-222.
- Frankena, W. K. Ethics (2nd ed.). Englewood Cliffs, N.J.: Prentice-Hall, 1973.
- Fromm, Erich. The art of loving. New York: Harper, 1956.
- Galbraith, R. E., and Jones, T. M. Teaching strategies for moral dilemmas. Social Education, 1975, 39, 16-22.
- Galbraith, R. E., and Jones, T. M. Moral reasoning: A teaching handbook for adapting Kohlberg to the classroom. Minneapolis: Greenhaven, 1976.
- Hall, Robert T., and Davis, John U. Moral education: In theory and practice. Buffalo, N.Y.: Prometheus, 1975.
- Harris, Richard J. A primer of multivariate statistics. New York: Academic, 1975.
- Kemp, C. G. Changes in patterns of personal values in relation to open-closed belief systems. Unpublished doctoral dissertation, Michigan State University, 1957.
- Kohlberg, L. Moral development. In D. L. Sills (Ed.), International encyclopedia of the social sciences. Vol. 10. New York: Macmillan, 1968.
- Kohlberg, L. From is to ought: How to commit the naturalistic fallacy and get away with it in the study of moral development. In T. Mischel (Ed.), Cognitive development and epistemology. New York: Academic, 1971. (a)
- Kohlberg, L. Stages of moral development as a basis for moral education. In C. M. Beck, B. S. Crittenden, and E. V. Sullivan, Moral education. Toronto: University of Toronto Press, 1971. (b)
- Kohlberg, L. The contribution of developmental psychology to education--examples from moral education. Educational Psychologist, 1973, 10, 2-14.
- Kohlberg, L. The cognitive-developmental approach to moral education. Phi Delta Kappan, 1975, 56, 670-677.
- Kohlberg, L., Colby, A., Gibbs, J., Speicher-Dubin, B., and Power, C. Assessing moral stages: A manual. Cambridge, Mass.: Harvard University, 1976.

- Kohlberg, L., and Kramer, R. Continuities and discontinuities in childhood and adult moral development. Human Development, 1969, 12, 93-120.
- Kohlberg, L., and Turiel, E. Moral development and moral education. In G. S. Lesser (Ed.), Psychology and educational practice. Glenview, Ill.: Scott, Foresman, 1971.
- Krathwohl, D. R., Bloom, B. S., and Masia, B. B. Taxonomy of educational objectives. Handbook II: Affective domain. New York: McKay, 1964.
- Kurtines, W., and Greif, E. B. The development of moral thought: Review and evaluation of Kohlberg's approach. Psychological Bulletin, 1974, 81, 453-470.
- Lawrence, Jeannette A. Moral judgment intervention studies using the Defining Issues Test. Unpublished manuscript, University of Minnesota, 1977.
- McDougal, William. An introduction to social psychology (33rd ed.). London: Methuen, 1950 (1908).
- McGeorge, C. The susceptibility to faking of the Defining Issues Test of moral development. Developmental Psychology, 1975, 11, 108.
- McNemar, Quinn. Psychological statistics (3rd ed.). New York: Wiley, 1962.
- Mischel, W. Toward a cognitive social learning reconceptualization of personality. Psychological Review, 1973, 80, 252-283.
- Neill, A. S. Summerhill. New York: Hart, 1960.
- Nie, N. H., Hull, C. N., Jenkins, J. G., Steinbrenner, K., and Bent, D. H. SPSS. Statistical package for the social sciences (2nd ed.). New York: McGraw-Hill, 1975.
- Panowitsch, H. Change and stability in the Defining Issues Test. Unpublished doctoral dissertation, University of Minnesota, 1975.
- Piaget, J. The moral judgment of the child (2nd ed.). Glencoe, Ill.: Free Press, 1948.
- Rawls, John. A theory of justice. Cambridge, Mass.: Harvard University Press, 1971.
- Rest, J. Developmental psychology as a guide to value education: A review of "Kohlbergian" programs. Review of Educational Research, 1974, 44, 241-259. (a)

- Rest, J. Manual for the Defining Issues Test. University of Minnesota, 1974. (b)
- Rest, J. Moral judgment related to sample characteristics. Final report to the National Institute of Mental Health, Grant No. RO3 MH 24988. University of Minnesota, 1976.
- Rest, J., Cooper, D., Coder, R., Masanz, J., and Anderson, D. Judging the important issues in moral dilemmas--An objective measure of development. Developmental Psychology, 1974, 10, 491-501.
- Rokeach, M. The open and closed mind. New York: Basic Books, 1960.
- Simpson, E. L. Moral development research: A case study of scientific cultural bias. Human Development, 1974, 17, 81-106.
- Skinner, B. F. Walden Two. New York: Macmillan, 1976.
- Swartzendruber, L. A test item file to accompany Kagan and Havemann's Psychology: An introduction (2nd ed.). New York: Harcourt Brace Jovanovich, 1974.
- Turiel, E. An experimental test of the sequentiality of developmental stages in the child's moral judgments. Journal of Personality and Social Psychology, 1966, 3, 611-618.
- Turiel, E. Developmental processes in the child's moral thinking. In P. Massen, J. Langer, and M. Covington (Eds.), Trends and issues in developmental psychology. New York: Holt, Rinehart and Winston, 1969.
- Watson, Goodwin, and Glaser, E. M. Critical Thinking Appraisal. New York: Harcourt Brace Jovanovich, 1964.
- Wilson, John. Practical methods of moral education. London: Heinemann, 1972.



## APPENDIX A

### THE ORIGINAL SIX-STAGE COGNITIVE-DEVELOPMENTAL MODEL OF MORAL REASONING PROPOSED BY KOHLBERG (1973)

#### I. PRECONVENTIONAL LEVEL

At this level the child is responsive to cultural rules and labels of good and bad, right or wrong, but interprets these labels in terms of either the physical or the hedonistic consequences of action (punishment, reward, exchange of favors) or in terms of the physical power of those who enunciate the rules and labels. The level is divided into the following two stages:

Stage 1: The punishment and obedience orientation. The physical consequences of action determine its goodness or badness regardless of the human meaning or value of these consequences. Avoidance of punishment and unquestioning deference to power are valued in their own right, not in terms of respect for an underlying moral order supported by punishment and authority (the latter being Stage 4).

Stage 2: The instrumental relativist orientation. Right action consists of that which instrumentally satisfies one's own needs and occasionally the needs of others. Human relations are viewed in terms like those of the market place. Elements of fairness, of reciprocity and equal sharing are present, but they are always interpreted in a physical pragmatic way. Reciprocity is a matter of "you scratch my back and I'll scratch yours," not of loyalty, gratitude or justice.

#### II. CONVENTIONAL LEVEL

At this level, maintaining the expectations of the individual's family, group, or nation is perceived as valuable in its own right, regardless of immediate and obvious consequences. The attitude is not only one of conformity to personal expectations and social order, but of loyalty to it, of actively

maintaining, supporting, and justifying the order and of identifying with the persons or group involved in it. At this level, there are the following two stages:

Stage 3: The interpersonal concordance or "good boy-nice girl" orientation. Good behavior is that which pleases or helps others and is approved by them. There is much conformity to stereotypical images of what is majority or "natural" behavior. Behavior is frequently judged by intention--"he means well" becomes important for the first time. One earns approval by being "nice."

Stage 4: The "law and order" orientation. There is orientation toward authority, fixed rules, and the maintenance of the social order. Right behavior consists of doing one's duty, showing respect for authority and maintaining the given social order for its own sake.

### III. POST-CONVENTIONAL, AUTONOMOUS, OR PRINCIPLED LEVEL.

At this level, there is a clear effort to define moral values and principles which have validity and application apart from the authority of the groups or persons holding these principles and apart from the individual's own identification with these groups. This level again has two stages:

Stage 5: The social-contract legalistic orientation generally with utilitarian overtones. Right action tends to be defined in terms of general individual rights and in terms of standards which have been critically examined and agreed upon by the whole society. There is a clear awareness of the relativism of personal values and opinions and a corresponding emphasis upon procedural rules for reaching consensus. Aside from what is constitutionally and democratically agreed upon, the right is a matter of personal "values" and "opinion." The result is an emphasis upon the "legal point of view," but with an emphasis upon the possibility of changing law in terms of rational considerations of social utility, (rather than freezing it in terms of Stage 4 "law and order"). Outside the legal realm, free agreement, and contract is the binding element of obligation. This is the "official" morality of the American government and Constitution.

Stage 6: The universal ethical principle orientation.

Right is defined by the decision of conscience in accord with self-chosen ethical principles appealing to logical comprehensiveness, universality, and consistency. These principles are abstract and ethical (the Golden Rule, the categorical imperative) they are not concrete moral rules like the Ten Commandments. At heart, these are universal principles of justice, of the reciprocity and equality of the human rights and of respect for the dignity of human beings as individual persons.

## APPENDIX B

### COURSE SYLLABUS: A CHARTER FOR SKIPPER, CADRE, AND CREW

Welcome aboard! Let's liken this class to a submarine departing from the Portsmouth Shipyard in search of the exotic waters of psychology. Yet doubts could arise. Might not the inevitable drowsiness of late afternoon learning soon pile up? Like an overload of ballast, could it not sink us to the bottom?

Not necessarily. You see, there was method in the madness of deliberately scheduling this course for an unpopular time slot. Only then are enough small rooms available to supplement lectures with small group discussions led by undergraduate teaching assistants (TAs). Hopefully, discussion topics like Ms. Phyllis Schlafly's views on homemakers, Ms. Anita Bryant's grasp of God's word, or Gov. Meldrim Thomson's scientific expertise on pot will keep all of you awake. And afloat.

In coordinating discussions and other activities to keep a large class awake, it is almost a necessity to run a tight ship. Hopefully this document will help us do that by specifying what we all can legitimately expect of each other. Thus, it is a charter for the skipper, the cadre, and the crew.

#### BILL OF FARE

A Look at the Undercurrents Before Launching Ship*.....	71
What's This About a Doctoral Dissertation?*	71
Logistics: Dull but Crucial.....	72
Required Books and Readings: Packing Your Gear.....	74
Study Time: A No-Nonsense Matter.....	74
The Ultimate Destination: One of Two Connected Seaports.....	74
The Grading System: Your Pay for Pitching in.....	75
Methods A and B: Different On-the-Job Training Programs.....	79

\*If you are in a hurry, you can skip these two parts now.

## A LOOK AT THE UNDERCURRENTS BEFORE LAUNCHING SHIP

I should give you fair warning about several of my biases, which will, like deep undercurrents, push us and pull us throughout the voyage. First, within the broad spectrum of colorful topics that we will view, I regard social psych and its complementary field of personality as particularly brilliant in their shared potential for enhancing the dignity of the human race. Second, although education requires remembering the printed and spoken word, I strongly suspect that a far more consequential component in a truly powerful education is the promotion of active thinking skills. Thus, the first and second currents merge to call for constructive thinking about social problems. For this purpose the one educational tool that seems to stand head and shoulders above the rest is the small group discussion, especially when it promotes active engagement in controversial issues.

Finally, as a third undercurrent, we must recognize research as the basis for all of psychology. Only through a specific type of research, namely experimentation, can the personal biases expressed above be put to the acid test. The course will indeed try to do just that. It will be a doctoral dissertation.

## WHAT'S THIS ABOUT A DOCTORAL DISSERTATION?

Both you and I--as student/subject and instructor/researcher--will be partaking in a doctoral dissertation on group discussions of controversial issues. You will receive laboratory credit for your part in completing some questionnaires. Also, from both my periodic explanations and your participation in a real life experiment you will learn some key ins and outs of the scientific method applied to psychology.

Specifically, the experimental stakes rest on two comparisons among students. First, the research will compare you with students in other 401 sections, none of which use TAs and regular group discussions. Second, after you have been randomly and impartially assigned to either of two highly

promising group discussion methods--called Methods A and B-- the experiment will compare those of you under one method with those under the other. Thus, the first comparison is between different 401 sections, whereas the second is within only your own.

But what will you be compared on? Good question. Although the answer to this could run for pages, it boils down to two primary sets of measures. First, there are your own evaluations of the course. Second, there are your own attitudes about social issues.

Attitudes? But wait a minute! Will someone try to indoctrinate you? The answer to this one is that even if any of us were to try, it probably would not work, at least not on anyone of college caliber. Instead, we will simply use discussions to provide you with the opportunity to develop attitudes along with intellect. By constantly comparing and contrasting your views with those of others who are also well informed and capable, you can mutually enlighten each other. Sometimes you can offer enlightenment to them, and hopefully sometimes they can offer enlightenment to you. The crucial factor in determining attitudinal growth, of course, will be your own definition of "enlightenment", and the extent to which you are open to it if indeed some of it comes your way. Hopefully, others will also be open to the insights that you offer to them.

Before getting too far adrift, however, let us turn from these more general matters to a more specific breakdown of the course as it is plotted.

#### LOGISTICS: DULL BUT CRUCIAL

Like remembering to pack a toothbrush, attention to details can prevent careless decay--perhaps in one's grade. The following items specify class and group meeting places and times, all on Tuesdays and Thursdays, as well as any private meetings you may want with me.

Meetings: Lecture-Only Days (Sep 8, 13, 15)

Hamilton Smith 127-128, 3:40-5:00

Meetings: Lecture/Discussion Days

For half of students:

Lecture in Ham Smith in 127-128, 3:40-4:30

Group meetings in separate rooms, 4:35-5:25

For other half of students:

Group meetings in separate rooms, 3:40-4:30

Lecture in Ham Smith 127-128, 4:35-5:25

Immediately before class on Tue 20 Sep a posted announcement will assign students to times, groups, and rooms

Meetings: Exam Days

If the last digit of your Social Security No. is odd:

Ham Smith 127-128, 3:40-5:00

If the last digit is even:

Ham Smith 216-217, 3:40-5:00

My Office, Phone, and Hours

Hamilton Smith 12B (862-2360): Wed 2:00-4:00, Fri 12:00-2:00. Also, I can talk at appropriate times by appointment. Please do not be shy. Your tuition fully entitles you to service, and I often benefit from talking personally with students.

Number of contact hours

You may have noted that on all but a few days your contact time (that is, time in lectures or groups) runs until 5:25 instead of 5:00, the later stated in the time and room schedule of the University. This altered arrangement is necessary to allow the minimum of 50 minutes for lectures or groups. However--and this point is quite important--your total number of contact hours for this course will be approximately the same as for any other section of Psych 401. Among the compensating

reasons are these: unlike the other sections, your four-hour laboratory requirement is included in your class time; there is no final exam; and on some days there will be no lecture (so that I can attend a few group meetings). Thus, you will average about three contact hours per week, as in other sections.

Now let's return to more interesting matters, such as where we are going and then how we will get there.

#### REQUIRED BOOKS AND READINGS: PACKING YOUR GEAR

The required literary gear for the trip is listed at the end of the Calendar. The total cost will be about \$27 (plus over \$4 for an optional but widely used study guide). These costs may seem steep, but as consolation, you will not have to set foot in the library or spend money and time doing any xeroxing.

#### STUDY TIME: A NO-NONSENSE MATTER

As a freshman--nearly everyone in this class is--it is vital to get started on the right foot. That means, for one thing, allotting sufficient time to each course. You will find that for this one the Calendar has been designed to require very nearly the same amount of work each week, namely, about ten to twelve hours worth, including contact time. Stated in career terms this four-credit course in a normal sixteen-credit load will require what is about one fourth of a salaried worker's time.

If any unfortunate souls saddled with heavy outside commitments are looking for a light, easy course, they have a right to know now during the add/drop period that this course is about the worst possible one to select. I cannot be too emphatic about this point.

#### THE ULTIMATE DESTINATION: ONE OF TWO CONNECTED SEAPORTS

The goals of the course are two destinations, but only one is ultimate. They are seaports located side by side. The first, Knowsport, is a refuge for those who ploddingly memorize a host of ideas about psych. The second, Thinksport, is a mecca



for the more adventuresome, who have made memorization fun by concomitantly getting high looking for relations among the ideas and finding applications to real life as it could and indeed should be.

More specifically, let me spell out the goals:

1. Knowing is the passive ability to report what was heard or read, in this case about psychological facts, principles, and theories.
2. Thinking is the active ability to interrelate and apply what is known, in this case about psych to real people and situations.

You are, of course, free to choose whatever destination you please. By choosing only the first you would pass the course but shortchange yourself. By choosing the second you would automatically be choosing the first also. Furthermore, I am sure you have what it takes to do the second, namely, to see relations within statements such as this one: Knowing is to Thinking as low grades are to high ones, as dull jobs are to fulfilling ones, as puny paychecks are to \_\_\_\_\_ . You supply the rest, thereby proving my point about which goal is for you.

THE GRADING SYSTEM: YOUR PAY FOR PITCHING IN

The more you accomplish the two preceding goals, particularly the second, the more pay you will get at the Registrar's Office. Your grade will be translated from a maximum of 1000 Course Points (CPs). This total results from group discussions (300 CPs), Exam One (250 CPs), Exam Two (200 CPs), a term paper (150 CPs), and a laboratory requirement (100 CPs).

#### Group Discussions (300 CPs)

The main goal here is Thinking. Since it will often result in the expression of opinions we will guarantee impartiality in grading it by using only these three unbiased, all-or-none criteria for each discussion meeting:

1. Attending (5 CPs or none)--being there or not.
2. Talking (5 CPs or none)--saying a bare minimum or not.
3. Writing (10 CPs or none)--turning in a short, minimally acceptable written statement or not.

The written statements will be described in more detail when you are assigned to your group. To receive credit for one you must actually attend the discussion to which the statement pertains. For example, for the three criteria, in order, 0 CPs + 0 CPs + 10 CPs = 10 CPs is illegal. However, 5 CPs + 5 CPs + 0 CPs is legal, although neglectful.

Only the 18 group meetings for which there is a specific paperback assignment are eligible for grading (except for the meeting on Oct 6). You can earn credit in as many as 15 of these meetings, thereby allowing you 3 excused absences. Additional absences will not be excused unless documented evidence can be provided for all absences, including the 3 absences that have already been used up. In other words, it is unwise to waste your 3 excused absences. You may need them in a pinch.

#### Exam One (250 CPs)

The goals of this exercise are both Knowing and Thinking in roughly equal parts. The exam will ask multiple-choice and essay questions--each type aimed at both skills, believe it or not. Each of the two paperbacks--Summerhill and Walden Two--will be worth 50 CPs, thereby allowing 150 CPs for the text and other readings.

Students who miss Exam One will be excused only by presenting documented justification. In such cases a cumulative make-up exam--that is, one over the entire course--will then be substituted on Dec 23, the time scheduled by the University.

#### Exam Two (200 CPs)

Again the goals are Knowing and Thinking, both tapped by questions like those on Exam One. Only the material assigned

after Exam One is fair game. The one paperback--The Art of Loving--will be worth 50 CPs, thereby allowing 150 CPs for the text and other readings.

Students who miss Exam Two will fall within the same make-up provisions as those who miss Exam One.

### The Term Paper (150 CPs)

The goal of Thinking is primary here, and with it you can really clean up. However, Knowing will also earn some credit.

Start by reading the four personality theories assigned for Weeks 10-12. Extract what you view to be the most promising ideas from at least two of these four major views. Then integrate these notions harmoniously into an eclectic personality theory that is exclusively your own. Only TYPEWRITTEN Reports will be accepted, and four to six pages is about right.

Your term paper will be graded independently by two TAs. Only the higher of the two grades will count. Thus, your ideas will get much closer scrutiny than I alone could give. To insure grading on merit only, I am asking that you identify the paper only by Social Security Number.

### The Laboratory Requirement (100 CPs)

This requirement, calling for Knowing and Thinking, enables you to see firsthand how psychologists get their knowledge and where it comes from. Similar requirements upon college students across our nation have furnished much--perhaps most--of the findings in your text by Kagan & Havemann.

At UNH the Dept of Psych requires four hours outside of class. However, since the research in this course is occurring within the class itself, we can kill two birds with one stone. In class I am asking you to spend four separate hours completing questionnaires about this research. From time to time in lectures, I will explain the basics of what we are doing. Finally, after completing your last questionnaire on Dec 15 I will hand you a more illuminating summary.

Then integrate your experiences, my oral explanations, and the printed summary into an essay or lab report that could

explain the experiment to any person on the street. She might want to see--and I am requiring--both the following topics and some elaboration on each:

- General area of psych
- Statement of problem
- Specific hypothesis tested
- Variables
  - Independent
  - Dependent
- Control procedures used
- Implications of the research
- Subjective evaluation:
  - Clarity
  - Value
  - Feelings

Please be sure to include a full paragraph of perfectly frank subjective evaluation--whether positive or negative. This can often help a researcher immensely in seeing things he or she might otherwise overlook.

The report must be TYPEWRITTEN on about two pages. It will be your very last assignment. I will personally read it and award either 0 or 100 CPs, with no in-betweens. Nearly everyone usually gets full credit.

#### Converting Course Points into Your Grade

Upon summing your CPs at the end of the course, I will convert them to a letter grade by combining two standards. First, according to the Registrar, the average instructor awards about 20 percent of the students A's, 40 percent B's, 30 percent C's, and 10 percent D's; conspicuously low students, if any, get F's. Second, although any notable overall performance differences between this class and my previous ones are unlikely, such gaps could modify the previous percentages somewhat in either direction.

Let's kill a possible misconception about the relative importance of your five CP subtotals, whose maximums are 300, 250, 200, 150, and 100 CPs, as you may recall. Their real contribution to your grade does not necessarily go in that order. Instead, the most important subtotal is the one on which you and your classmates spread out the most among yourselves. The greatest spread will probably occur on the second subtotal, Exam One (maximum 250 CPs), on which most of the scores will probably hover from about 175 CPs to 225 CPs, thereby making it easy to assign different grades to different students. For this reason, the most important parts of your grade will probably be the two exams and the term paper. (However, you would seriously hurt yourself by failing to turn in, say, the Lab Report, because nearly everyone else will, thereby being 100 CPs higher and nearly all bunched together.)

#### The Administrative F: A Ghastly Scar

If for any reason you fall hopelessly behind during the early part of the course, then please remedy your plight: drop the course officially at Thompson Hall by Tue 25 Oct. Failure to do so has a tragic consequence: an administrative F.

#### METHODS A AND B: DIFFERENT ON-THE-JOB TRAINING PROGRAMS

The group discussions can be regarded as training programs that help us get to our ultimate destination, Thinksport. As indicated earlier, two such programs will be used: Method A and Method B. You will be randomly assigned to only one. Random, of course, means that your assignment is made by the "dice" of research--not by you or by me. Although I will describe your assigned method to you in more detail later, perhaps you would like a brief sketch of both now.

#### Method A

Here the first half of each group meeting is devoted to the discussion of a social dilemma. For example, Does a Radcliffe student have the right to start dating others if her present lover's psychologist warns her that doing so will certainly

cause him to jump off of the Prudential Center? The last half of the meeting is devoted to a similar dilemma based on the current paperback reading assignment. After the meeting each student writes a short summary of the core ideas that some other group member expressed.

#### Method B

In this case each student prepares for the group meeting by writing a short discussion question. For instance, Why do you like or dislike the childrearing policy in Walden Two, where child training specialists spend more time with children than parents do? Then at the meeting itself each student is assigned roughly equal time to entertain a short discussion on his or her question.

#### Seeing Both Methods in Action

Both of these methods will soon be demonstrated to you in lecture by specialists. The TAs, in whom I have immense confidence, will play the role of students while I act as the TA. Hopefully it will get the semester off to a lively start. It did last Fall.

## APPENDIX C

### GROUP DISCUSSIONS: METHOD A

Showing much promise, this method is issue centered. It stresses two activities, both aimed at the goal of Thinking. First, you talk over a psychological issue with students like yourself. Second, you listen quite carefully to each of them to facilitate comparing their views with yours. The issues are taken from both your daily paperback assignment and from a set of psychologically oriented issues, The Fireside Fifty, which will be handed out.

Let us now crystalize these broad strokes into a more explicit, standard operating procedure (SOP) for each afternoon discussion.

1. One week prior to the discussion your group and TA will select one of three issues described quite briefly in The Fireside Fifty. Then, during the week glance at the issue in its alternate, expanded form, which is also in The Fireside Fifty. However, even the expanded form will not contain sufficient information for a reasonable person to make an informed judgment about it. Therefore, simply entertain some tentative thoughts on it to share with your colleagues, who will also share theirs with you.
2. At the discussion meeting a week later you will be asked to discuss your views on, first, the previously identified issue and, second, a new issue suggested by the current paperback assignment. Hopefully on each issue there will be a healthy disagreement, which your TA will try to promote, because without it an issue never gets off the ground, and discussion can become labored.

Discussions will stress not which side of an issue you lean toward, but rather the reasons for this stance.

3. A crucial aspect of Method A is genuine listening, as embodied in the following requirement. At the start of each meeting your TA will--unknown to you-- designate one student as the Key Person. This student's identity, known not even to him- or herself, will not be revealed until the very end of the meeting. After the meeting you are to summarize what the Key Person said about both issues. Do this on a Listening Sheet.
4. At the next meeting submit your Listening Sheet original to your TA, who will award it full credit unless you showed little recollection of what the Key Person said. Submit the copy to the Key Person, who is then responsible for writing either "Accepted" or "Rejected" on it plus any comments that he or she may want to add. However, neither the Key Person's assessment nor comments--if any-- will affect your grade. He or she must return this copy to you at the next group meeting.

Your academic pay for discussion contributions is specified in the Course Syllabus, pages 75 through 76. This part of the course should elicit no sweat from anyone who reads the assignments and listens to others. Indeed, much optimism has already been expressed about it.



APPENDIX D

SOCIAL ISSUES FOR GROUP DISCUSSIONS UNDER METHOD A:

THE FIRESIDE FIFTY

Just as Presidents Roosevelt and Carter held fireside chats about social issues, so can you in group meetings. The following fifty issues for use under Method A are, of course, more psychological and interpersonal than political and international. Further, issues that interlock directly with the current course topic for the scheduled discussion date are marked by asterisks (\*); only then is the topic listed.

Although the first issue, called Captain's Dilemma, and the second one, titled Voluntary Euthanasia, are required discussion material, there is choice among the remaining forty-eight. Specifically, for each date only one of the three listed issues is to be chosen by your group. (Credit goes to the Kohlberg staff for the first issue, to Hall and Davis, 1975, for the second, and to Blatt, Colby, and Speicher, 1974, for the rest.)

<u>Date</u>	<u>Topic</u>	<u>Social Issues</u>
Sep 22	-----	<u>Captain's Dilemma</u> : Should a captain on an under-supplied lifeboat order anyone overboard so that more might live?
Sep 27	-----	<u>Voluntary Euthanasia</u> : Should a terminally ill father add a few months to his life by hospitalization, if its steep expense costs his wife and teenage daughters their home and college education?
Sep 29	-----	<u>Honor System</u> : On a take-home exam to be graded on a curve, should you help a friend who asks for answers? <u>Shoplifting</u> : In a supermarket should Ms. Jackson report a young man who is slipping food into his coat? <u>Yellow Fever</u> : Should doctors have injected yellow fever into unknowing volunteers in the hope of finding a cure to save many worker's lives, as well as building the Panama Canal?

<u>Date</u>	<u>Topic</u>	<u>Social Issues</u>
Oct 4	-----	<p><u>Sexual Exploitation:</u> If a girl will have intercourse only with someone who loves her, is it wrong for a guy to seek and have sex with her by insincerely acting serious?</p> <p><u>Sterilization:</u> If a man fails to support the many children he produces in and out of wedlock, should a court require that he choose between jail and sterilization (which is a minor operation)?</p> <p><u>Husbandless Impregnation:</u> If a wife wants to give birth to children, not adopt them, but her husband cannot produce, should she try to get pregnant with another man?</p>
Oct 6	-----	No issues today.
Oct 11	Physio- logical Psych	<p>*<u>Homemade LSD:</u> Is it wrong for a chemistry student to produce LSD in her own home, provided that no one else gets it?</p> <p>*<u>Drugs: Soft vs. Hard:</u> If Joe's brother is selling both soft and hard drugs and the latter is ruining people's lives, should Joe report him?</p> <p>*<u>Heroin and Life Sentences:</u> Should people who repeatedly sell heroin get mandatory life sentences to prevent the drug's ruinous effects on others?</p>
Oct 13	Emotion	<p>*<u>Reclaiming One's Child:</u> If an unmarried mother's illegitimate and only child is adopted, and she later gets married but cannot give birth, should she seek to reclaim the child she now longs for?</p> <p>*<u>Parents vs. Boyfriend:</u> If an adolescent girl is confusing passion with an irreversibly serious relation with her boyfriend, should her parents try to stop it?</p> <p><u>Marriage by False Advertising:</u> If a girl lures a guy into marriage by falsely claiming to be wealthy, but after the marriage he learns the truth, does he have grounds for divorce?</p>
Oct 18	Motivation (& Women's Issues)	* <u>Daughter vs. Son:</u> Should parents be more strict about the sexual relations of their adolescent daughter than about the ones of her twin brother?

<u>Date</u>	<u>Topic</u>	<u>Social Issues</u>
		<p><u>*Affair for Money:</u> Should a woman become her boss' mistress to pay for her husband's cancer treatments and provide for her family?</p> <p><u>*Accidental Injury:</u> If, with attention seeking motives, a boy removes a girl's chair but accidentally breaks her back, should her parents sue his?</p>
Oct 20	Self Actuation	<p><u>*Career vs. Family:</u> When offered an exciting job requiring much travel, should a father take it or continue to devote lots of time and affection to his family?</p> <p><u>Protecting a Mate:</u> If two thugs with guns convincingly order a guy to abandon his girl or pay with his life, then is his honor or the threat to his life more important?</p> <p><u>Sex as a Need:</u> If the father in a family with four children becomes paralyzed and the mother still needs sex, should she have an extramarital affair or get a divorce?</p>
Oct 25	Review	<p><u>*Sex Roles in Marriage:</u> If two students want to get married but realize that the cost will force one of them out of school for now, should it be the guy or the girl?</p> <p><u>*Saving Lives by Torture:</u> In order to obtain crucial information about drug rings that are ruining lives, should law officials use physical means to get information from convicted ring members?</p> <p><u>*Marrying for Money:</u> If an aging, wealthy widower needs companionship, should a young woman with an eye on his will provide it by marrying him?</p>
Oct 27	-----	No issues today.
Nov 1	-----	No issues today.
Nov 3	Develop- mental Psych (from Nov 1)	<p><u>*Forced Busing:</u> If busing is the only feasible way to achieve integration, yet a parent is convinced that it hurts his own child's education, is he justified in trying to stop it?</p> <p><u>*Doctor vs. Parents:</u> Should a doctor save a dying child without the permission of the parents--who are Christian Scientists--or the courts?</p>

<u>Date</u>	<u>Topic</u>	<u>Social Issues</u>
		<u>Spouse Swapping</u> : If sex has become dull for two middle aged couples, but all four partners agree to "swap" for intercourse, should they?
Nov 8	-----	<u>Living Together</u> : If a couple want to try living together as a trial for marriage, should they?  <u>Reproducing Retardates</u> : If a retarded couple are highly likely to have only retarded children, who burden society heavily, should this reproduction be allowed?
		<u>Health Care</u> : Should doctors be forced onto government payrolls to insure that people unable to pay, particularly their children, are guaranteed adequate health care?
Nov 10	Social	<u>*Anti-Semitism</u> : If banks are reluctant to hire and promote Jewish people, should a Jewish applicant for bank teller lie about his nationality to correct for the bias?  <u>*Discrimination in Clubs</u> : Does a private country club have the right to exclude from its membership any race, religion, or sex?  <u>*Discrimination in Real Estate</u> : To prevent a drop in their own house values, should white neighbors insist that other neighbors sell their houses only to whites?
Nov. 15	Aggression	<u>*Killer Turned Priest</u> : If a German officer had obeyed Nazi orders to kill many innocent Poles but had refused to kill still other ones, and he later becomes a priest, should he at any time be tried in court as a war criminal?  <u>*My Lai Massacre</u> : At My Lai, Vietnam, where Lt. Calley ordered his company to kill all the villagers, including children, was Sgt. Bernhart, the only soldier who refused, more right or less right than every other American there?  <u>*Gun Control</u> : Should violent crimes be reduced by restricting dangerous weapons or should people have the right to use them in self defense?

<u>Date</u>	<u>Topic</u>	<u>Social Issues</u>
Nov 17	Freud	<p><u>*Extramarital Accident:</u> Should Ms. Hamilton, a wife and mother who got pregnant in a secret affair with her boss, tell her husband who the real father probably is if the husband would not otherwise know?</p> <p><u>*Paralyzed Husband:</u> If a young husband becomes paralyzed before his wife has conceived any children, whom she wants badly, should they get divorced?</p> <p><u>*Lonely Wife:</u> If newlyweds have no children, and the husband is away at war, and the wife is quite lonely, should she have an affair with someone until her husband comes home?</p>
Nov 22	-----	<p><u>Robin Hood:</u> If a stingy old man with an inherited fortune plays cards with a shrewd young man who gives all winnings to people in need, is it right for the young man to cheat?</p> <p><u>Astronauts in Trouble:</u> If a space capsule containing four astronauts will probably explode unless one of them sacrifices himself to fix the problem outside, should the captain do it himself or order someone else to do it?</p> <p><u>Support of Parents:</u> Should aging parents be supported by their children or by the government?</p>
Nov. 29	-----	<p><u>Homosexual Tenants:</u> If some apartment tenants object that other tenants are homosexuals and can prove it, thus threatening the building's reputation, should the landlord ask the homosexuals to move out?</p> <p><u>Unwilling Kidney Donor:</u> If only one dying man has the perfect tissue match for a kidney transplant that will save a girl's life, but he wants his dead body to remain intact, should the doctor perform the transplant anyway?</p> <p><u>Draft Resistance:</u> If people do not feel that their country is right in a foreign war, should they refuse to be drafted?</p>
Dec 1	-----	<p><u>Plagiarism:</u> In courses that grade on the curve and, thus, reward undetected plagiarism only at the expense of more honest students, is an automatic F in the course too severe a penalty for this offense?</p>

<u>Date</u>	<u>Topic</u>	<u>Social Issues</u>
Dec 6	Review	<p data-bbox="607 374 1351 527"><u>Mother's Life vs. Child's:</u> If either the mother's life or the child's can be saved in the delivery room, and the husband says he wants the child, should the doctor comply?</p> <p data-bbox="607 540 1351 693"><u>Bomb Shelter:</u> If the Jones' bomb shelter has enough air for only themselves, but the Smiths are trying to break in, should Mr. Jones shoot them as a last resort to keep them out?</p> <p data-bbox="607 706 1351 859"><u>*Premarital Sex:</u> If a girl cannot explain why she feels that premarital sex is wrong, but still feels strongly about it, should she adopt her boyfriend's rationale that it is right?</p> <p data-bbox="607 872 1351 995"><u>*Dating Between Races:</u> Should a white girl date a black guy if their relation could lead to marriage and children, whom some people might reject?</p> <p data-bbox="607 1008 1351 1157"><u>*Dating Between Religions:</u> Should a Catholic girl marry a Jewish guy if her truly devout parents maintain that her act will destroy their lives, especially if any grandchildren become Jewish?</p>

## APPENDIX E

### GROUP DISCUSSIONS: METHOD B

Highly popular last year, this method is student centered. It stresses two activities, both aimed at the goal of Thinking. First, from one of the paperbacks you select a topic of interest to students like yourself. Second, you discuss the matter with others.

The more the topic integrates the paperback with the text and real life, the more it will aid both your understanding of them and your resulting grade on exams. Thus, there are three ingredients to be mixed--when feasible--by your own ingenuity: the current assignment in the paperback, anything from the text, and whatever is relevant in your own life or that of others. Also, any topic that compares and contrasts two paperbacks, such as Summerhill vs Walden Two, say, on educating children, is marvelous. However, do not be dismayed if you cannot integrate multiple sources, for that is not always possible.

Let us now crystalize these broad strokes into a more explicit, standard operating procedure (SOP) for each afternoon discussion.

1. Prior to the discussion select a topic--the more controversial the better--from the appropriate paperback assignment, which is usually about 50 pages, sometimes less. Then express it on a Question Card.
2. At the discussion meeting your TA will call upon you to read your question to your colleagues to stimulate discussion for about ten minutes. Also, it is you who must then try to keep discussion going in a productive direction. This responsibility will also be shared by your TA when he or she deems it appropriate.

3. Upon leaving your meeting, submit your Question Card to your TA. She or he will award it full credit, unless the question showed only a tiny amount of thinking and put most people to sleep.

Your academic pay for discussion contributions is on a reasonable, noncompetitive, all-or-none basis. Please see the Course Syllabus, pages 5 through 6 for details. (Note that it refers to your Question Card as a "short, minimally acceptable written statement.") This part of the course is rarely any sweat for anyone who reads the assignments. In fact, last year most students said that an earlier, less refined version of Method B was what made the course, which they rated highly.



## APPENDIX F

### EXEMPLARY DISCUSSION QUESTIONS WRITTEN BY STUDENTS

#### UNDER METHOD B

Here is a sample of some successful questions written by students. (Slight editing has improved economy in wording.)

#### From Summerhill

1. In what way is the adjustment from high school life to college life similar to the adjustment that transfer students to Summerhill had to make?
2. Neill's criterion for a successful individual is "the ability to work joyfully and to live positively." In our own society what personal qualities or material possessions must you obtain to feel successful and satisfied?
3. Everyone in Summerhill has the right to vote, and each vote carries the same weight. What benefits do you think there are to this system for the children? Are there any shortcomings?
4. Neill claims that free children are not influenced easily. Are his children really similar or different from Walden Two kids? Are either unlike the sheep who are controlled by a piece of string, that is, are either really free?

#### From Walden Two

1. Would a society like Walden Two be acceptable to most Americans today? Why or why not?
2. Ms. Meyerson said, "many of our women manage to appear quite beautiful because they are not required to dress within strict limits." Does our present society have any limits on dressing, or are people of all ages free to dress as they please, regardless of changing styles?

3. In Walden Two work is kept at a minimum. If applied to our own society Frazier's method would reduce the 40-hour week sharply. Would it work or not?
4. Are the Walden Two children, who are kept in separate buildings from their parents, harmed or benefitted from this experience?