TRAINING OF MEDICAL STUDENTS IN A PROBLEM - SOLVING SKILL: THE GENERATION OF DIAGNOSTIC PROBLEM FORMULATIONS

> Dissortation for the Degree of Ph. D. MICHIGAN STATE UNIVERSITY LINDA KATHLEEN ALLAL 1973

NAT 21 A GAL

#### 1 - F + 3

м. . .

• ••

•

۸. . .

.

,

• • •

.

.

3 1293 10024 0674

4 - • •

• • •

LIBRARY \* licbigan State University

1

.

.







### ABSTRACT

### TRAINING OF MEDICAL STUDENTS IN A PROBLEM-SOLVING SKILL: THE GENERATION OF DIAGNOSTIC PROBLEM FORMULATIONS

By

Linda Kathleen Allal

The purpose of this research was to develop and test experimentally a procedure for training second-year medical students in one aspect of medical problem solving: namely, the generation of diagnostic problem formulations on the basis of cues obtained during the initial minutes of the doctor-patient encounter. Recent investigations of medical problem solving (e.g., Elstein, <u>et al.</u>, 1972) have found that the early generation of problem formulations (or hypotheses) is a major feature of the experienced physician's cognitive activity in conducting a clinical workup. Training which explicitly focuses on this process may be expected to improve the medical student's ability to generate appropriate early problem formulations, and, thereby, aid him in making the transition from classroom to clinical practice.

The training model developed and tested in this study included two major components: (1) having the student practice the task of generating initial problem formulations under conditions which simulate the early part of the clinical

encounter, and (2) providing the student with feedback based on the performance of this task by a sample of experienced physicians. Color films which present a "physician's eye view" of the first four-six minutes in a doctor-patient encounter were used to simulate the conditions of the early part of the workup. Two type of feedback materials were utilized: (1) "outcome feedback," consisting of written materials summarizing the outcomes (i.e., problem formulations generated and cues associated with each) characteristic of the experienced physicians who viewed each film, and (2) "process feedback," consisting of written materials and a second set of films designed to portray the processes by which the experienced physicians arrived at their problem formulations outcomes. The process feedback films consisted of a second version of each of the original films in which "think aloud" recordings, interposed at various points in the dialogue, were used to portray the processes typically going on inside the physician's head as he observes and interviews the patient.

A training experiment was designed to test the following hypotheses: (1) that the training model would significantly improve second-year medical students' skill in generating diagnostic problem formulations, and (2) that the model would be significantly more effective when it provides both outcome and process feedback than when it provides outcome feedback only.

In preparation for the training experiment eight color films were produced and data on problem formulation outcomes and processes were collected from a sample of eight experienced physicians. On the basis of these data, the feedback materials and posttest scoring keys were prepared. The training experiment included three conditions: two treatment conditions and a posttest-only control condition, with 16 second-year medical students randomly assigned to each condition. Both treatment conditions involved the application of two-component training model, but they differed with respect to the type of feedback that was provided: under one condition, the subject received outcome feedback only; under the second condition, the subject received both outcome and process feedback. Training consisted of three weekly sessions, followed by a posttest session on the fourth week. The subject's posttest performance was evaluated by means of four dependent variables: (1) a problem formulation score, (2) a cue utilization score, (3) a classification of cues with respect to problem formulations score, and (4) a relationships among problem formulations score.

Analysis of covariance on the posttest data yielded the following results. Hypothesis 1 was supported: a significant difference was found between the trained groups and the control group on the dependent variable of major interest (problem formulation score), although not on the other variables. Moreover, supplemental analyses indicated the trained groups' problem formulation performance was superior on both quantitative and qualitative dimensions. Hypothesis 2 was not supported: there were no significant differences between the two trained groups on any of the dependent variables. Several explanations for the ineffectiveness of the process feedback were given consideration, in particular the possibility that the subjects in the group which received outcome feedback only were able, on the basis of this material, to generate their own process feedback.

In addition to the analysis of the results of the training experiment, the data collected from the sample of experienced physicians were analysed in detail. On the basis of this analysis several tentative conclusions were drawn. First, with respect to problem formulation outcomes, it was found that the result of the physician's informationprocessing activity during the early part of the workup is not a unidimensional list of problem formulations, but a structured set of problem formulations which may be described in terms of four features: (1) hierarchical organization, (2) competing formulations, (3) multiple subspaces and (4) functional relationships. Secondly, with respect to problem formulation processes, it was found that direct associative retrieval, rather than strategy-guided search, appears to be the primary cognitive mechanism involved. Discussion of the implications of the results of the training experiment--for future research and instructional development--dealt with the following topics: (1) extensions of the training model to other types of medical problemsolving skills; (2) use of other types of media (e.g., slidetape units) to simulate the early part of the workup; (3) questions pertaining to the feedback component of the model; (4) instructional applications of the training materials. The findings from the analysis of the physician data were also discussed with respect to their implications for future research on medical problem solving.

### References

Elstein, A. S.; Kagan, N.; Shulman, L. S.; Jason, H.; and Loupe, M. J. Methods and theory in the study of medical inquiry. Journal of Medical Education, 1972, 47, 85-92.

# TRAINING OF MEDICAL STUDENTS IN A PROBLEM-SOLVING

### SKILL: THE GENERATION OF DIAGNOSTIC

### PROBLEM FORMULATIONS

By

Linda Kathleen Allal

### A DISSERTATION

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

### DOCTOR OF PHILOSOPHY

Department of Counseling, Personnel Services and Educational Psychology

### ACKNOWLEDGMENTS

Appreciation is extended to the members of my dissertation committee, Professor Arthur S. Elstein, Professor Andrew C. Porter, Professor Rose Zacks, and, in particular, to my committee chairman, Professor Lee S. Shulman, for the able guidance, the encouragement and the perceptive criticism which they offered throughout the course of this research.

A special debt of gratitude is owed to Donald Gragg, M.D., David N. Ostrow, M.D., and Michael Spooner, M.D., for their extensive aid in the preparation of the research materials. I am also indebted to Gregory Loftus for his assistance in the administration of the training sessions, and to Mary Fedewa for her typing of the research materials and the many drafts of the dissertation.

This research was supported in part by National Institute of Health Grants PM-00041 and 71-2208.

ii

TABLE OF CONTENTS

e

																Page
LIST	OF TAP	BLES	••	•	•	•	•	. •	•	•	•	•	•	•	•	vi
LIST	OF FIC	GURES	•	•	•	•	•	•	•	•	•	•	•	•	•	ix
DEFIN	ITION	OF T	ERMS	•	•	•	•	•	•	•	•	•	•	•	•	x
Chapt	er															
I.	INTRO	ODUCT	ION	•	•	•	•	•	•	•	•	•	•	•	٠	1
		Over	view Rati The Desi Rese	of ona Tra gn arcl	the le inir of f h Hy	St ng the /pc	Mode Study	el idy ses	• • •	• • • •	• • •	• • •	• • •	• • •	• • •	4 8 9 12 14
II.	REVII LITEI	EW OF RATUR	THEC E .	RET:	ICAI •	: I •	SSUI •	es •	AND •	RES •	•	кСН •	•	•	•	16
		Medi	cal P	rob	lem	Sc	lvir	ng	•	•	•	•	•	•	•	16
		Init Prob	ial P lem S	rob olv	lem ing	Fc •	ormu: •	lat •	ions •	5 in •	n Ma	edio •	cal •	•	•	26
		Deve	lopme The The	nt o Sim Fee	of t ulat dbao	the tic ck	e Tra on Co Comp	ain omp oon	ing onen ent	Moo nt o of	del of the	the e Mo	Mod Dde:	del 1	• •	35 35 40
III.	METH	DD	•••	•	•	•	•	•	•	•	•	•	•	•	•	44
		Prod	uctio	on o	f tł	ne	Filr	ns	•	•	•	•	•	•	•	44
		Coll	ectic Samp Mate Proc Anal Trair	on o ole ria edu ysi	f Pl ls re s Exp	nys • • •	sicia	an • • •	Data with	a · ·	• • •	• • • •	Yea:	• • •	• • •	50 50 53 55 57
		Medi	cal S	tud	ent	S	•	•	•	•	•	•	•	•	•	58

Page

	Experimental Procedure	. 59 . 64 . 67 . 71 . 72 . 76 . 78 . 83 . 89 . 91 . 93 . 103 . 107 . 108 . 108
IV.	ANALYSIS OF THE PHYSICIAN DATA	. 110
	Generation of the First Problem Formulation	. 111
	The Structure of a Set of Initial Problem Formulations	. 113 . 114 . 123 . 127
	Processes Involved in Generating Initial Problem Formulations	. 127 . 148
V.	RESULTS OF THE TRAINING EXPERIMENT	. 150
	Tests of Experimental Hypotheses Reliability Estimates Results of Hypothesis Tests Relationships among Dependent	<ul> <li>150</li> <li>150</li> <li>154</li> </ul>
	Supplemental Analyses	• 100 171
	Results of Additional Posttest Tasks Treatment-Control Differences in	. 171
	Problem Formulations	. 177 . 186
VI.	CONCLUSIONS AND IMPLICATIONS	. 200
	Problem Formulation Outcomes and Processes in the Experienced Physician	. 200

		Training of Medical Students in the	
•		Formulations	204 204 205 212
	REFER	NCES	214
	APPEN	ICES	221
	Α.	Case Outline for Film 1: "A 21-year-old	
		college senior"	222
	В.	Process Checklist	226
	с.	Training Materials	<b>2</b> 30
	D.	Additional Posttest Tasks	264
	E.	Questionnaire	<b>2</b> 69
	F.	Posttest Scoring Keys and Scoring Instructions .	275
	G	Analysis of the Structure of a Set of Problem	
	0.	Formulations	288
	H.	Additional Data	290
	I.	Homogeneity of Regression	294
	J.	Modifications of the Outcome Feedback Version	301
		or the fragming materials	20T

Page

L

### LIST OF TABLES

Table			Page
1.	The Eight Films	•	46
2.	Characteristics of the Physician Sample	•	52
3.	The Experimental Procedure	•	62
4.	Properties of the Feedback Presented Under the Two Treatment Conditions	•	70
5.	Selected Characteristics of the Student Sample	•	89
6.	Results of the Pilot Test	•	91
7.	Features Characteristic of Individual Sets of Problem Formulations, by Film and by Subject	•	. 118
8.	Number of Problem Formulations, and Number of Subspaces: Average and Range by Film, and by Subject	•	124
9.	The Relative Frequency with which each Process Checklist Item was Checked	•	135
10.	Classification of Checklist Items on Two Dimensions: (A) Degree of Subject Stability, and (B) Degree of Task Stability	•	137
11.	Inter-scorer Reliability Coefficients on the Variables CUE, PF, CUE-PF and R-PF	•	151
12.	Generalizability Coefficients, and Within-Group Correlation Coefficients (Between Tasks) on the Dependent Variables CUE, PF, CUE-PF and R-PF	•	154
13.	Means and Standard Deviations on the Dependent Variables and Covariate, by Experimental Condition	•	157
14.	Adjusted Means on the Dependent Variables, by Experimental Condition	•	157

#### Table

Page

### Table

31.	Scheffé Post Hoc Comparisons on Number of Subspaces and Number of Problem	
	Formulations	293
32.	Regression and Correlation Coefficients for the Covariate with each Dependent Variable, by	
	Experimental Condition	298
33.	Tests of Homogeneity of Regression for CUE, PF,	200
	$CUE-PF, R-PF \cdot \cdot$	299
34.	Regression and Correlation Coefficients for Treatment II with One Deviant Subject	
	Eliminated	300

### LIST OF FIGURES

.

.

Figur	e	Page
1.	A sample of the Outcome Feedback for Film 1	73
2.	Relationship Between Cognitive Outcomes and the Dependent Variable Scores CUE, PF and CUE-PF .	95
3.	Structural Diagram of the Composite Set of Problem Formulations Generated by the Physician Sample for Film 1	116
4.	Bivariate Distribution of Treatment II Scores on the Covariate and the Dependent Variable CUE .	300

#### DEFINITION OF TERMS

<u>CUE</u>--A cue is an element of data pertaining to the patient's physical and/or psycho-social condition which the physician utilizes to generate diagnostic problem formulations. The cues of particular concern in this study are those obtained during the early part of the clinical workup: namely, (1) symptoms of a physical and/or psycho-social nature reported by the patient, and (2) nonverbal cues, including physical signs of illness and general psychosocial indices, observed by the physician.

PROBLEM FORMULATION--A problem formulation is a label, having potential diagnostic and/or management implications, which the physician generates on the basis of cues obtained from the workup. It represents a tentative "working diagnosis," or hypothesis, which can account for some portion of the cues obtained. A problem formulation may range from the highly general (e.g., "organic disorder," "psychological problem") to the highly specific (e.g., "myocardial infarction," "glomerulonephritis") depending on the adequacy Of the cues at hand.

<u>WORKUP</u>--This term designates a clinical encounter between a physician and a patient. A workup typically includes: (1) an interview of the patient to elicit medical,

х

personal and family history, (2) physical examination of the patient, and (3) ancillary diagnostic procedures (e.g., laboratory tests). A workup may take place under a variety of conditions, with respect to setting (e.g., private office, hospital outpatient or inpatient facilities), number of previous encounters between the patient and doctor, and nature of the patient's medical problem (e.g., routine check-up, present illness, medical emergency). The type of workup with which the present study is concerned is the office visit in which the physician encounters a patient for the first time and is presented with one or more complaints pertaining to a present illness.

### CHAPTER I

### INTRODUCTION

The past two decades have seen considerable ferment and innovation in the field of medical education. The traditional medical school curriculum, organized in terms of discipline-centered coursework in the basic sciences during the first half of the program, followed by supervised clinical work (clerkships) during the second half of the program, has come under substantial criticism as a method of preparing students for their future responsibilities as clinicians. Although criticism has covered a wide range of issues, much of it has focused on two concerns. The first is that the medical student be introduced to his future role as a clinician at a much earlier point in his studies than has traditionally been the case. This concern has led to the inclusion of courses dealing with clinical skills during the first half of the medical school program, and to attempts to integrate the teaching of the basic science disciplines with orientation to the clinician's role. A second concern has been that the student's introduction to his role as a clinician focus on patients and their problems, rather than on abstract principles of medical science, as has been characteristic of the traditional curriculum. One

of the earliest responses to this concern was the organization of clinical clerkships around the principle of "comprehensive medical care," with a new emphasis on the psychological and sociological dimensions of medical problems as manifested in the individual patient (Hammond and Kern, 1959; Reader and Goss, 1967). More recently, this concern has led to efforts to devise new instructional methods: (a) to develop the student's interpersonal skills, needed for effective and humane interaction with the patient, and (b) to develop the student's information-processing skills, needed for effective medical diagnosis and decisionmaking. Methods to achieve the first goal have included the creation of courses on "doctor-patient relations" and training in interpersonal techniques. With respect to the second goal, efforts have been directed toward the development of what may be termed "problem-solving" approaches to the teaching of clinical skills. These efforts have included the introduction, early in the program, of coursework dealing with "focal problems" (Ways et al., 1973), the abandonment of the traditional "case presentation" method of clinical teaching in favor of an approach emphasizing the generation and testing of diagnostic hypotheses (Engel, 1971), and the instruction of students in the use of "problem-oriented" methods of medical record keeping (Weed, 1969; Ways et al., 1972).

"Problem-solving" approaches in medical education, like their counterparts in the curriculum reforms of the

1950s and 60s in primary and secondary education, exhibit considerable diversity of objectives and methods. There are two premises, however, which are common to all such approaches. The first premise is that the practice of clinical medicine is in essence a problem-solving activity (Miller, 1962; Elstein, et al., 1972; Barrows and Bennett, 1972; Ways, et al., 1973). Thus, development of the information-processing skills necessary for medical problem solving is considered to be a primary curriculum objective. The second premise is that the teaching of these skills can best be achieved by means of a problem-solving mode of instruction: (a) by providing the student with opportunities early in his program to practice medical problem solving under conditions which closely approximate the conditions of clinical practice (Ways, et al., 1973), and (b) by designing clerkship experiences which will encourage the student to approach the clinical workup as a problem-solving task (Miller, 1962; Engel, 1971). Both of these premises rest, implicitly, on psychological and educational principles of long standing and respected parentage: most notably the conception of human cognitive activity and of the educational process advocated by Dewey (1963, 1938) and Bruner (1960, 1966). However, their application in the context of medical education still represents a major challenge at the present date.

### Overview of the Study

The physician's activity in a clinical setting is exceedingly complex. It involves both cognitive informationprocessing skills and affective interpersonal skills, and, in terms of outcomes, requires both diagnostic judgments and management decisions. The present study, like the larger research project<sup>1</sup> of which it is a part, has limited the scope of its investigation to: (a) the physician's <u>cognitive</u> skills, and (b) the way in which he uses these skills to arrive at a <u>diagnosis</u>. Moreover, the present study focuses on only a very small, but highly important, portion of the diagnostic process: namely, the physician's informationprocessing activities during the first five minutes of his encounter with a patient.

In recent years there have been a number of studies (Elstein, et al., 1972; Barrows and Bennett, 1972) designed to investigate in depth the cognitive activities of the experienced physician in carrying out a clinical workup. The outstanding feature of these studies, as compared to earlier research efforts (e.g., Kleinmuntz, 1968; Rimoldi, 1963), is their use of simulated patients in order to study

<sup>&</sup>lt;sup>1</sup>This study is one of several conducted by the Medical Inquiry Project, directed by Arthur Elstein and Lee Shulman, Office of Medical Education, Research and Development, Michigan State University. Reports on other studies conducted by the project are found in Elstein, et al. (1972), Gordon (1973) and Sprafka (1973). A comprehensive final report on the activities of the project is forthcoming in January 1974.

the physician's activity in a naturalistic setting, closely resembling the conditions of actual clinical practice, and yet achieve an adequate degree of experimental control. Probably the most salient feature of medical problem solving to emerge from these investigations is the critical role of hypotheses, generated by the physician during the earliest minutes of his encounter with a patient, and subsequently tested by the collection of data during the remainder of the workup. Other investigators (Wortman, 1972; Schwartz and Simon, 1970) have also identified hypothesis generation as a fundamental step in the diagnostic process. The notion that medical problem solving begins with the generation of diagnostic hypotheses is hardly novel in one sense: theories V of problem solving, both classical (Dewey, 1938) and current (Shulman et al., 1968; Newell and Simon, 1972) have posited that the generation of some form of conceptual framework (whether in the form of hypotheses, problem formulations or a problem space) is a major early step in attempting to solve complex problems. However, the notion is quite novel within the context of clinical medicine, where the diagnostic process has long been considered and taught, in the classical empiricist tradition, as an essentially inductive procedure in which thoroughness and objectivity of data collection should precede **Consideration** of diagnostic possibilities (e.g., Harvey and Bordley, 1970). Thus, a major import of recent research is the finding that experienced physicians, despite their

training in an inductive, "reserve judgment" approach to diagnosis, almost universally employ an approach characterized by early hypothesis generation and subsequent hypothesis testing.

The purpose of the present study, broadly stated, is: (a) to develop, and (b) to test experimentally a procedure for training medical students in the generation of diagnostic problem formulations based on the data obtained during the earliest minutes of a clinical workup. In this study the term "problem formulation" (rather than "hypothesis") has been used to refer to the outcomes of the physician's information-processing activity during the early part of the workup. This term has been chosen because it is one with which the particular medical student population participating in the experiment was accustomed through their coursework on the use of problem-oriented methods of medical record keeping. Conceptually, however, it has the same meaning as the term "hypothesis," as employed in the research by Elstein, et al., Barrows and others. In focusing on the first five minutes of the clinical encounter, this study is concerned with two types of information-processing skill: (1) the detection and encoding of cues (i.e., elements of data having diagnostic relevance) presented by the patient during the early minutes of the workup, and (2) the use of these cues to generate an initial set of diagnostic problem formulations. Obviously, the process of testing these initial

formulations by means of further data collection is of crucial importance in the resolution of diagnostic problems. The present study, however, limits its investigation to the training of students in the process of <u>generating</u> these initial formulations.

Before presenting an overview of the study, one more preliminary comment is in order. It should be recognized that the adoption of a "problem-solving" approach to training medical students does not necessarily imply training in the early generation of diagnostic problem formulations. Some current methods of clinical training do emphasize the generation of hypotheses early in the workup (e.g., Engel, 1971). But others do not. For example, the steps in medical problem solving, as defined by Ways, et al. (1973, p. 566) include: "(1) sensing that a problem exists; (2) collecting the data (from history, physical examination, and ancillary diagnostic methods) . . .; (3) rationally defining and formulating the problem(s); (4) deciding upon a course of action. . . " In this sequence, it will be noted, the act of generating problem formulations (step 3) is at the end of the workup (step 2), which is where the traditional "reserve judgment" view of the diagnostic process has always placed consideration of diagnostic possibilities. Thus, a major feature of the present experimental study is that it attempts to train medical students in the generation of diagnostic problem formulations at a very early point in the workup, a point

at which only a very small portion of potentially relevant data is available. As far as this particular goal is concerned, there have been very few instructional precedents, and virtually no research precedents.

### Rationale

The rationale for training medical students in the early generation of diagnostic problem formulations rests on two arguments:

1. A major conclusion to be drawn from much of the psychological literature on problem solving and inquiry (Dewey, 1938; Shulman <u>et al.</u>, 1968; Newell and Simon, 1972) is that, when confronted with any complex problematic situation, the human information-processor inevitably generates some sort of conceptual framework as an early step in his search for a solution. The Elstein, <u>et al.</u> (1972) and Barrows and Bennett (1972) investigations provide evidence that this conclusion applies to medical problem solving as well. Thus, training of medical students in the early generation of diagnostic problem formulations is consistent with what is known about the cognitive processes of human problem-solvers in general, and of experienced physicians in particular.

2. Since physicians, despite their training in the traditional "reserve judgment" approach to diagnosis, are found to generate problem formulations very early in the workup, it is to be anticipated that medical students, even without special training, would tend to approach

medical problems in a similar manner, and that this tendency would be reinforced by increasing clinical experience. However, it is believed that training, which explicitly focuses on this process, will improve the medical student's ability to generate <u>appropriate</u> early problem formulations, and, thereby, aid him in making the transition from classroom to clinical practice.

#### The Training Model

The training model developed and tested in the present study includes two major components: (1) having the student practice the task of generating initial problem formulations under conditions which <u>simulate</u> the early part of the clinical encounter, and (2) providing the student with <u>feedback</u> based on the performance of this task by experienced physicians.

The first component in the model, the use of simulation exercises, is based on the educational principle that problem-solving skills can best be taught by providing the student with opportunities to encounter and attempt to solve a range of problems which closely approximate, in breadth and complexity, the problems which he will encounter in the real world (Dewey, 1963; Bruner, 1966; Gagné, 1971). In the present case the student's encounter with a series of patients, having various medical complaints and diverse demographic characteristics, is simulated by means of color

films which present a "physician's eye view" of the first five minutes of a clinical workup.

The second component in the training model, the provision of feedback on the performance of the task by experienced physicians, is designed to enable the student to evaluate his own performance on a given exercise, and, across the full set of exercises, to increase his skill in attaining problem formulation outcomes similar to those of the experienced physician. Two types of feedback are employed in this experiment: (1) feedback on the outcomes of physicians' problem formulation activity during the earliest part of the workup; and (2) feedback on the processes by which physicians arrive at these outcomes. Both types of feedback are based on data obtained from a sample of experienced physicians who viewed the training films. The first type of feedback presents, in written form, the problem formulations (and cues associated with each) generated by the physicians, and is designed to indicate both the commonalities and the range of diversity that were found in their out-The second type of feedback includes a special comes. version of each of the training films in which "think aloud" recordings are interposed at various points in the dialogue. The purpose of these films is to provide the student with a simulated portrayal of the problem formulation processes typically going on inside the physician's head as he observes and interviews the patient. In addition, written materials

are used to summarize the similarities and differences among physicians with respect to processes of generating a set of initial problem formulations.

The feedback employed in this study has several rather special features which distinguish it from most traditional types of feedback. First, the notion of utililizing feedback to provide the learner with "process models" is obviously guite foreign to the behaviorist tradition of feedback. Second, the type of outcome feedback provided in this study is closer to what has been termed "cognitive" feedback (Hammond and Summers, 1972) than to the classical types of outcome feedback used in learning experiments or programmed instruction. A major feature of the feedback is that it does not provide the student with a single "correct" model of either outcomes or processes. Rather it indicates both the convergent and the divergent aspects of the performance of experienced physicians: i.e., the commonalities characteristic of nearly all physicians, as well as the ways in which they differ, with respect to the generation of a set of initial problem formulations. Thus, in utilizing the feedback to evaluate his own performance the student must engage in a series of relatively complex cognitive activities: he must examine, synthesize and draw inferences from a sample of the performances of experienced practitioners in his field.

### Design of the Study

The study involved both a developmental phase and an experimental phase. The developmental phase included: (1) production of a set of films of the first five minutes in eight doctor-patient encounters, and (2) collection of data on problem formulation outcomes and processes from a sample of eight experienced physicians. These data were then utilized as the basis for development of the training and evaluation materials employed in the experimental phase of the study.

The second phase of the study consisted of a training experiment involving 48 second-year medical students, randomly assigned to three conditions:

> Treatment I: Training with Outcome Feedback; Posttest

Treatment II: Training with Outcome and Process Feedback; Posttest

Control: Posttest only.

Both treatment conditions involved application of the "simulation exercises, plus feedback" training model. The conditions differed, however, with respect to the type of feedback provided. Under one condition (Treatment I) the subject was provided with outcome feedback only, while under the other condition (Treatment II) the subject received both

outcome and process feedback.<sup>1</sup> The third condition consisted of a posttest-only control group.

In designing the experiment it was recognized there are a number of researchable questions of potential interest with respect to the proposed training model, not all of which could be dealt with within the context of the present study. The decision as to which questions would be asked, and which would be deferred for subsequent investigation, was largely dictated by the author's conception of the research strategy that is most fruitful for the educational psychologist to follow in developing a new method of training. This strategy suggests the following research priorities: first, to determine the effectiveness of the best training "package" one can devise, as compared to the results already being attained by means of existing procedures; second, to investigate those manipulations of the package that are likely to have the greatest educational relevance; third, providing that some variation of the total package produces the desired results, to determine the effects of the separate components in the package. A major reason for the use of such a strategy is that the components of a training model may, in combination, produce the desired effect even though any single component would have an insignificant effect. A second reason for the use

<sup>&</sup>lt;sup>1</sup>It should be noted that a "process feedback only" condition is not possible: feedback pertaining to the processes of problem formulation requires that the outcomes of these processes be mentioned.

of such a strategy is that the application of experimental findings to on-going programs would indeed proceed at a snail's pace if it were to be predicated on the construction of new training models out of components whose separate effects had each been experimentally demonstrated. Tn terms of the present study, this strategy has been translated into an experiment that is designed: (1) to determine the effectiveness of the "simulation exercise, plus feedback" training model as a whole, compared to existing conditions (i.e., a posttest-only control), and (2) to determine the relative effectiveness of the model when its feedback component is manipulated so as to provide, in the one case, outcome feedback only, and, in the second case, outcome and process feedback. Other questions not dealt with in this study, but which should be investigated, include: (1) how cost-effective are films, as compared to either higherfidelity (e.g., simulated patients) or lower-fidelity (e.g., slide-tape combinations) means of simulating the early part of the doctor-patient encounter? (2) how effective are the simulation exercises without feedback of any type? (3) what variations of the model are most effective for students at different stages in the medical school curriculum?

#### Research Hypotheses

The major purpose of the study was to test experimentally the following hypotheses:

Hypothesis 1:

That a training model consisting of: (1) problemsolving exercises in which films are used to simulate the conditions of the early part of the clinical workup, and (2) feedback based on data from a sample of experienced physicians, will significantly improve second-year medical students' skill in the generation of an initial set of diagnostic problem formulations.

Hypothesis 2:

That the training model will be significantly more effective when it provides both outcome and process feedback, than when it provides outcome feedback only.

A secondary purpose of the study was to specify, in greater detail than has been forthcoming from previous investigations, the nature of the problem formulation component of medical problem solving. This purpose was accomplished by an analysis of the physician data, designed to address the following questions:

- How early in the clinical workup does the physician begin to generate problem formulations?
- 2. What is the structure of a set of initial problem formulations?
- 3. What cognitive processes are involved in the generation of initial problem formulations?

#### CHAPTER II

## REVIEW OF THEORETICAL ISSUES AND RESEARCH LITERATURE

The purpose of this chapter is to examine theoretical issues and research literature of relevance to three topics: (1) medical problem solving, (2) initial problem formulations in medical problem solving, (3) the development of a model for training medical students in the generation of initial problem formulations.

### Medical Problem Solving

In attempting to define the nature of medical problem solving a logical starting point is to examine the viewpoint of the medical profession itself. A description of the approach to diagnosis--which is held to constitute good clinical practice and taught to medical students--can be derived from a review of several well-known, authoritative volumes in the field. In Harvey and Bordley's <u>Differential</u> <u>Diagnosis</u> (1970, p. 7) and in <u>The Principles and Practice</u> <u>of Medicine</u> (18th edition), edited by Harvey, <u>et al.</u> (1972, p. 39), the process by which the physician arrives at a diagnosis is described in terms of the following sequence of steps:

Steps in Diagnosis

- 1. Collecting the Facts
  - a. Clinical history.
  - b. Physical examination.
  - c. Ancillary examinations.
  - d. Observation of the course of the illness.
- 2. Analyzing the Facts
  - a. Critically evaluate the collected data.
  - b. List reliable findings in order of apparent importance.
  - c. Select one or preferably two or three central features.
  - d. List diseases in which these central features are encountered.
  - e. Reach final diagnosis by selecting from the listed diseases either: (1) the single disease which best explains all the facts, or, if this is not possible, (2) the several diseases each of which best explains some of the facts.
  - Review all the evidence--both positive and negative--with the final diagnosis in mind.

<u>Harrison's Principles of Internal Medicine</u> (6th edition, Wintrobe, <u>et al.</u>, 1970) suggests a similar sequence, proceeding from collection of clinical data to evaluation of data with respect to diagnostic possibilities. In each of these volumes it is noted that "working diagnoses" (or hypotheses) may emerge at various points in the workup, and that to some degree the physician engages in data analysis concurrently with data collection. However, neither of these activities is considered to be a central feature of the diagnostic process. In discussing the physician's data collection activity over the course of the workup major emphasis is placed on thoroughness and objectivity, particularly with respect to history taking and physical examination (e.g., "the patient must be literally
scrutinized from top to bottom in an objective search for abnormalities," <u>Harrison's Principles</u>, p. 5). In discussing the process of data analysis, it is noted that a preliminary evaluation of history and physical data enables the physician to make a judicious selection of ancillary (laboratory) examinations, but, in largest part, the evaluation of data with respect to diagnostic possibilities is to take place at the end of the workup, after the physician has completed his search for the facts of the case. In sum, the physician is "to begin, as in all scientific research, by marshalling all the facts, then proceeding with an unprejudiced analysis of the facts, and ending with the logical conclusion" (Harvey and Bordley, 1970, p. 3).

This view of the diagnostic process is based on a conception of scientific inquiry that: (1) is primarily inductive in nature, i.e., conclusions emerge out of an objective analysis of "all the facts," and (2) makes the classical empiricist assumption that autonomous facts (observables which exist independently of the observer and are objectively verifiable) constitute the ultimate arbiter of scientific claims of knowledge (theories, hypothesis or conclusions). This conception of scientific inquiry has been challenged by an impressive number of scholars of the philosophy of science. Kessel (1969), in his review of these scholars' viewpoints, argues that it is necessary to reject the empiricist notion of the autonomy and

objectivity of facts. He asserts that "the scientist's premises and presuppositions can and do play a significant role at all levels of his endeavor" (p. 1004). Moreover, he suggests, the safeguard of disciplined scientific inquiry does not lie in the scientist's attempting to banish these premises and presuppositions (as, it may be noted, the physician is admonished to  $do^1$ ), rather it lies: (1) in the scientist's making explicit (to himself and others) the premises which he is entertaining, and (2) in his "actively seeking to invent alternatives" to his current premises in order to insure that the success of his endeavor does not rest on the spurious ground that potentially refuting facts were never sought, or were misinterpreted (p. 1002). This latter recommendation calls to mind Chamberlin's classic article (1890; reprinted in Science, 1965) in which he argued that only by employing a method of "Multiple Working Hypotheses" can the scientist avoid the pitfall of biased data collection and interpretation.

A major challenge to the classical empiricist conception may be found in John Dewey's Logic: The Theory of <u>Inquiry</u> (1938). In this work Dewey combines both philosophical and psychological considerations in his development of a

<sup>&</sup>lt;sup>1</sup>In <u>Harrison's Principles</u> (1970, p. 6) it is stated that "the physician does not start with an open mind. . ., but with one prejudiced from knowledge of recent cases;" consequently, "he must struggle constantly to avoid the bias" that would interfere with the objective conduct of the clinical inquiry.

model of inquiry which coincides quite closely with the viewpoints summarized in Kessel (1969). First of all, Dewey asserts that the problem-solver's initial conceptualization of the problem, as well as the "ideas" (or hypotheses) that emerge from this conceptualization, play a crucial, "operational" role in the selection and interpretation of facts, and in the organization of facts into a coherent whole. Secondly, Dewey argues in favor of a dialectical view of inquiry in which ideas are as much arbiters of facts as facts are of ideas.

The orders of fact, which present themselves in consequence of the experimental observations the ideas call out and direct, are trial facts. They are provisional. They are 'facts' if they are observed by sound organs and techniques. But they are not on that account the facts of the case. They are tested or 'proved' with respect to their evidential function just as much as ideas (hypotheses) are tested with reference to their power to exercise the function of resolution (1938, p. 114).

A further challenge to the classical empiricist conception of the inquiry process is provided by contemporary research within the framework of "cognitive" or "informationprocessing" theories of psychology. There is a very sizable body of research findings, comprehensively reviewed in Neisser (1967), which suggest that a person's expectations (presuppositions, ideas or hypotheses) have a significant impact on even the simpler forms of cognitive activity, e.g., visual and auditory perception, visual and auditory memory. Research into higher-order cognitive processes,

such as problem solving (Newell and Simon, 1972), indicates that the internal representation of the task (which the person generates) initiates, organizes and directs his subsequent information-processing activities.

Although certain types of problem solving have been quite intensively investigated during the past several decades, the activities of the physician in attempting to solve diagnostic problems have been studied for a relatively brief number of years (Rimoldi, 1963; Kleinmuntz, 1968; Elstein, et al., 1972; Barrows and Bennett, 1972; Wortman, 1972; Schwartz and Simon, 1970). The theoretical orientation of most of this research derives from the cognitive or information-processing conceptions of human psychology which emerged in the late 1950s, as embodied in the work of Bruner, Goodnow and Austin (1956), Miller, Galanter and Pribram (1960), and Newell, Shaw and Simon (1958). Although investigations of medical problem solving have employed diverse types of diagnostic tasks, e.g., card sorting (Rimoldi, 1963), a variant of the game Twenty-Questions (Kleinmuntz, 1968), simulated clinical encounters involving actors trained to play the role of patients, i.e., "simulated patients" (Elstein, et al., 1972; Barrows and Bennett, 1972), most have utilized these tasks in the context of the methodological approach developed by Newell, Shaw and Simon (1958). A protocol of the physician's behavior--e.g., the questions he asks, maneuvers he performs--in conducting the diagnostic

task is an important source of data for describing some parameters of his activity, but description of the cognitive processes underlying his behavior relies primarily on introspective data obtained by having the subject "think aloud" as he attempts to solve the problem, and, additionally, in the case of the Elstein, <u>et al.</u> (1972) investigation, by a "stimulated recall" technique.

Research on medical problem solving has, thus far, (1) in the identification of several of the resulted: major features of the diagnostic process, and (2) in the specification--at a fairly general, descriptive level--of the sequence of events involved in this process. Present research efforts are still quite far, however, from achieving one of the major goals of the informationprocessing theorist: namely, the specification of cognitive mechanisms in terms of a computer program which simulates the physician's problem-solving activity. The findings which have been fairly well established by this research may be summarized as follows. The major characteristics of the physician's information-processing activity in conducting a workup are: (1) the early generation of multiple diagnostic hypotheses, and (2) the testing (revision and refinement) of these hypotheses by means of subsequent data collection. There may, depending on the difficulty of

<sup>&</sup>lt;sup>1</sup>An initial attempt at computer simulation of diagnostic problem solving has been undertaken by Wortman (1972).

1,

the case, be several iterations of the hypothesis generation/ hypothesis testing sequence. There are individual differences in <u>how</u> early a physician begins to generate hypotheses, but, nearly all physicians do so at a very early point (at least within the first five minutes of the workup).<sup>1</sup> Although the physician sometimes uses standard formats of data collection (e.g., asks routine history questions, performs the physical examination in a systematic head-to-toe manner), his mode of cognitive processing (i.e., of organizing, sorting, interpreting and synthesizing these data) is structured by the hypotheses he is entertaining.

Newell and Simon (1972) have proposed that the essence of an information-processing theory of human problem solving can be summarized in four propositions:

- A few, and only a few, gross characteristics of the human IPS (information-processing system) are invariant over task and problem solver.
- 2. These characteristics are sufficient to determine that a task environment is represented (in the IPS) as a problem space, and that problem solving takes place in a problem space.
- 3. The structure of the task environment determines the possible structures of the problem space.
- The structure of the problem space determines the possible programs that can be used for problem solving. (p. 788)

<sup>&</sup>lt;sup>1</sup>Analysis (by the author) of a portion of the data from Elstein, <u>et al.</u> (1972) investigation indicated that the percentage of subjects who began generating hypothesis within the first five minutes of the workup was 95.2%, 81.8% and 100% for three simulated clinical encounters. The number of questions asked by the physician prior to generating his first hypothesis was, on the average, 13.9, 20.8 and 7.4 for the three simulations.

The remainder of this section will be devoted to consideration of the following question: What are the features of the task environment in clinical medicine that are responsible for the hypothesis-guided nature of medical problem solving?

Although theories of problem solving--both classical (Dewey, 1938) and current (Shulman, et al., 1968; Newell and Simon, 1972) -- have posited that the generation of some sort of internal representation of the task is a major early step in attempting to solve complex problems, this representation does not take the form of multiple hypotheses regarding potential end states in the types of problem solving that have been most intensively investigated in the recent psychological literature (e.g., cryptarithmetic problems, logic problems, chess playing). On the other hand, the generation of multiple hypotheses is characteristic of most forms of scientific inquiry, including medical inquiry. Borrowing a distinction made by Bartlett (1958) between reasoning in "open" versus "closed" systems, it may be useful to consider various types of problem solving as lying at different points along a continuum, with science (including clinical medicine) lying toward the "open system" pole of the continuum, and logic and mathematics toward the "closed system" pole. (Chess playing would probably be classified at some intermediate point). Given this conceptual framework, we will now consider two characteristics of the task environment of open systems, in particular clinical medicine, which

may account for the hypothesis-guided nature of problem solving in such systems.

A first characteristic of the task environment in open systems is the indeterminacy of the end state. In closed system problems, such as cryptarithmetic and logic problems, the end state to be reached is fully specified in advance (e.g., show the DONALD + GERALD = ROBERT; prove that Ll is equivalent to L2). In an open system, such as clinical medicine, not only is the end state (i.e., the diagnosis) unspecified, its general configuration may take many forms (e.g., the patient may have a single disease, several separate diseases, several related diseases). The generation of hypotheses enables the physician to define a set of potential end states toward which his problem-solving activity may be directed, and thus tentatively "closes" the system in which he is operating.

A second distinction between problem solving in open versus closed systems pertains to the nature of solution criteria. For problems in closed systems, such as cryptarithmetic and logic, there are well-defined <u>a priori</u> criteria-known to the problem solver--which he may apply in order to determine that an appropriate solution has been reached. In an open system such as clinical medicine there are no such criteria. The criteria that do exist are <u>a posteriori</u> (i.e., the patient responds to the treatment, or his condition worsens), and, in general, are ambiguous. Occasionally there

are pathonogmonic findings that conclusively substantiate the correctness of a diagnosis, but most often the relationship between the physician's state of knowledge (the data he has obtained) and the solution to the problem (the patient's disease) is a probablistic one. However, the use of multiple working hypotheses provides a pragmatic criterion for judging that an appropriate solution has been reached: i.e., the physician concludes that "Having generated and tested hypotheses X, Y and Z, the data collected tend to rule out hypotheses X and Y, but tend to support hypothesis Z. Therefore, I will consider Z as the most tenable diagnosis on which to base treatment."

## Initial Problem Formulations in Medical Problem Solving

This section will examine in somewhat greater detail the diagnostic problem formulations (or hypotheses) which the physician generates in the early minutes of the clinical workup.<sup>1</sup> First, the findings of recent research regarding initial problem formulations will be reviewed. Second, both the functional advantages and potential risks involved in the early generation of problem formulations will be considered.

<sup>&</sup>lt;sup>1</sup>For reasons discussed in Chapter I (p. 6), the term "problem formulation" (rather than "hypothesis") is generally employed in this research to refer to the diagnostic labels which the physician generates on the basis of cues obtained during the workup.

A number of recent investigations have found that physicians begin to generate problem formulations at a very early point in the workup. Data from the Elstein, <u>et al.</u> (1972) investigation indicate that this process nearly always occurs within the first five minutes of the physician's encounter with the patient (see footnote 1, p. 23). Barrows and Bennett (1972, p. 275) have observed that hypotheses literally "pop" into the head of the clinician "almost before the interview begins." Schwartz and Simon (1970) have also observed that from the moment of patient contact and presentation of chief complaint, the physician begins to generate hypotheses.

Although the earliness with which the physician generates problem formulations appears to be fairly well established, current research findings are as yet rather sketchy as to the characteristics of these formulations, and as to the types of cognitive mechanisms involved in generating them. One characteristic that has been investigated is the number of problem formulations a physician generates. As yet unpublished data from the Elstein, <u>et al.</u> (1972) investigation indicate that, for three simulated workups, the number of hypotheses generated was, on the average, 6.7, 7.0 and 4.2. Barrows and Bennett (1972) report that the neurologists in their study always generated at least three hypotheses, and sometimes as many as five hypotheses. Wortman (1972) found that students solving

diagnostic type problems (involving concrete objects having various properties), as well as a neurologist whose diagnostic behavior was studied, tended to select cues that were associated with three <u>or</u> seven categories (objects/diseases). In sum, the research to date tends to suggest that the number of hypotheses a physician generates ranges from three to seven, which, it may be noted, coincides with Mandler's (1967) proposition that human information-processers organize and store information in terms of 5 ± 2 categories.

11

A second characteristic of problem formulations which has received some attention in the research literature is level of specificity. Kleinmuntz (1968), in a study of neurologists which involved a medical variant of "Twenty-Questions," found that his subjects tended to follow a general-to-specific strategy of hypothesis generation: beginning with very general formulations moving toward increasingly specific formulations. In a more recent paper (Wortman and Kleinmuntz, undated) it is suggested that although the physician follows a general-to-specific search strategy he tends to begin at the most specific level that is possible--given the adequacy of available cues--in the hierarchy of diagnostic categories stored in long-term memory. Barrows and Bennett (1972), in a study of neurologists carrying out simulated workups, reached a conclusion similar to Kleinmuntz's: namely, that the physician's initial hypotheses are quite general, but are

progressively shaped into more specific entities by means of a "coning down" strategy. Elstein, et al. (1972), on the other hand, found in a preliminary analysis of the problemsolving behavior of internists (working up simulated patients) that initial problem formulations were frequently quite specific. Subsequent analysis of the Elstein, et al. data, in which the author assisted, has indicated that a physician's early problem formulations are often heterogeneous with respect to level of specificity: he may simultaneously entertain a general hypothesis (e.g., "psychogenic paralysis"; "viral infection") and a highly specific hypothesis that can be either a competitor to, or a subset of, the general hypothesis (e.g., "multiple sclerosis"; "infectious mononucleosis"). In sum, it is probable that the physician's initial problem formulations cannot be characterized as a uniform list of either very general, or very specific, diagnostic considerations.

Although the observations of most investigators indicate that the generation of problem formulations on the basis of cues obtained during the earliest minutes of the workup occurs in a very rapid and virtually automatic manner, very little is known as yet regarding the types of cognitive mechanisms that are involved in this process. Different researchers have made various suggestions, including associative retrieval mechanisms (Elstein, <u>et al.</u>, 1972), pattern recognition mechanisms (Lusted and Stahl, 1963), strategy-based search mechanisms (Kleinmuntz, 1968; Kleinmuntz

and Wortman, undated; Barrows and Bennett, 1972). To date, however, there has not been enough research focused on the physician's information-processing activities during the earliest part of the workup in order to specify in detail which of these mechanisms (or what combination of mechanisms) is involved in the generation of initial problem formulations.

The remaining portion of this section will consider the functional advantages and potential risks that are likely to be involved in the early generation of diagnostic problem formulations. It is probable that the set of initial problem formulations which the physician generates early in the workup serves a dual cognitive function: (1) it provides an organizational framework for the storage of data obtained during the workup; and, (2) it provides a conceptual framework that guides the physician's data processing activity during the workup. The "storage function" of problem formulations relates to the role of organization in human memory. In medical problem solving, as in nearly any relatively complex task, the amount of information which must be stored greatly exceeds the capacity of short-term memory, i.e., 7 ± 2 symbols (Miller, 1956). Moreover, information in short-term memory rapidly decays, or is deleted by the incoming flow of new information, unless it is processed (rehearsed, recoded) and transfered to long-term memory, which has working storage capacity that is potentially limitless (Broadbent, 1958). A sizable body of research

literature, notably Miller, (1956), Mandler, (1967), Tulving and Donaldson, (1972), indicates that the organization of elementary units of information into "chunks," or "categories," increases the holding capacity of short-term memory, and facilitates the processes of transfer to, and retrieval from, long-term memory. In medical problem solving, it is probable that the problem formulations which the physician generates constitute the organizational framework which permits storage of the very large amount of information that is obtained over the course of the workup. Although the storage function of diagnostic problem formulations has not yet been investigated in depth, the finding that physicians failed to recall data that were not associated with one of the hypotheses they were entertaining (reported by Kleinmuntz, 1968, and Barrows and Bennett, 1972) would lend some support to this assertion.

The second cognitive function of initial problem formulations--which may be termed the "guidance function"-relates to the concept of the "problem space" proposed by Newell and Simon (1972). These theorists assert that the task environment is represented internally as a problem space, and that the structure of this problem space determines the programs (i.e., information-processing activities) to be used in the search for a solution. Any problem space, as defined by Newell and Simon, includes of two types of components: (1) "elements," which are symbolic structures

representing states of knowledge about the task, and (2) "operators," which are processes (procedures, methods) for producing new states of knowledge from existing ones. In clinical medicine, as in other domains of problem solving, the potential size of the problem space is enormous: there are a vast number of elements (states of knowledge about the patient) that could be obtained, and an exceedingly large number of potential operators (interview questions, physical examination maneuvers, laboratory tests) for obtaining them. The early generation of diagnostic problem formulations would appear to be a major strategy that is used by the physician to determine the regions of the potential problem space which are most likely to yield a solution. In sum, a set of problem formulations can be considered to constitute the framework of the functional problem space in which the physician conducts his search for a diagnosis.

In addition to the cognitive functions served by early problem formulation, there are two practical considerations which would tend to foster this activity. One is the time constraints under which the practicing physician has to work. Selective data collection, guided by a set of problem formulations, may be the only feasible way to complete a workup within the amount of time typically allotted for each patient. Secondly, there are occasions when the physical or psychological well-being of the patient requires that a management decision be taken

right away, before completion of the data collection involved in a thorough workup. To make such decisions the physician would have to be operating in an "early problem formulation," rather than a "reserve judgment," mode of reasoning.

Although the early generation of problem formulations presents a number of advantages, both cognitive and practical, which justify its use, it also entails potential risks. The primary risk is that of premature closure: i.e., the possibility that the physician may fail to seek data that could disconfirm his initial problem formulations, or fail to revise these formulations in the face of contradictory evidence, and thus arrive at an incorrect diagnosis. The literature on problem solving indicates that the initial conceptualization of the problem may create a mental set (Einstellung) that prevents the subject from shifting to more appropriate formulations as his search for a solution progresses (Luchins and Luchins, 1950). In addition, there is evidence that the human problem solver shows a marked reluctance to eliminate or revise an initial hypothesis as long as there is some confirmatory evidence in its facor (Wason, 1968). This may manifest itself as a failure to seek evidence that could disconfirm an hypothesis, or as a failure to interpret properly negative evidence that is obtained. The phenomena of set and psychological bias toward confirmatory evidence pose particular risks for

medical problem solving. The wide range of possible data collection options in a medical workup, as well as the probabilistic character of clinical findings (leaving room for considerable latitude for interpretation on the part of the physician), make it possible that even an experienced clinician may become inadvertently wedded to an incorrect initial formulation due to biased data collection and/or biased data interpretation procedures.

There appear to be two means which the physician may use to avoid these problem-solving pitfalls. One pertains to the process of generating problem formulations. Premature closure becomes much less likely if the physician generates multiple problem formulations, all of which are supported in some degree by the initially available data. In this case it becomes necessary to seek out and critically interpret additional data that will permit determination of a differential diagnosis. Moreover, if care is taken to generate problem formulations at a level of specificity that is appropriate to the data at hand, this would tend to facilitate subsequent revision of initial formulations (i.e., broadening or narrowing of the problem space) in the light of new data. A second means of overcoming the risks of early problem formulation pertains to the process of data acquisition. As Elstein, et al. (1972) have proposed, a major justification for the use of routine procedures of history taking and physical examination is to reduce the likelihood

of biased data collection. Such procedures would help to insure that an adequate test of initial formulations will be made, and that there will be sufficient opportunity for tacts to emerge which could lead to revision of initial formulations and/or generation of additional formulations. A second possible rationale for routine procedures is that their application would require a minimal investment of cognitive effort on the part of the experienced physician, and would thus enable him to focus a greater part of his attention (and working memory space) on the processing (interpretation, organization, synthesis) of data, rather than on the activity of collecting data.

## Development of the Training Model

This section will examine considerations related to the development of a model for training medical students in the generation of initial problem formulations.

## The Simulation Component of the Model

One component of the training model consists of having the student practice the problem-solving skill to be attained--i.e., the generation of initial problem formulations--under conditions which simulate the early part of the clinical encounter. As Shulman (1970) has pointed out, the choice of a mode of instruction is not necessarily dictated by the nature of the outcome behaviors to be acquired. In principle, for example, it would be possible to teach skills

ę

in clinical problem solving by an expository (or didactic) approach, or to teach basic concepts in medical science by a problem-solving (or discovery) approach. For some educational outcomes, e.g., the learning of concepts, principles or rules, there is considerable debate among educational psychologists as to the most appropriate mode of instruction, with Bruner (1966) advocating use of a discovery approach for the teaching of these skills, and Gagné (1970) arguing in favor of carefully programmed expository methods. An expository approach has been used with some success in training physician's assistants to employ clinical algorithms (step-by-step instructions) for dealing with a very limited set of medical problems, i.e., eleven acute illnesses (Sox, et al., 1973). However, this type of (algorithmic) rule learning is obviously not appropriate for training future physicians who will be called upon to solve a wide range of often complex medical problems, a task which would require, as it does in other domains (e.g., cryptarithmetic, logic, chess), the application of higher-order problem-solving skills such as cognitive strategies or heuristics. With respect to these higher-order skills, there is considerably more consensus among educational psychologists regarding instructional method. For these skills we find that Gagné (1971) would advocate -- in practice, although not necessarily in principle--an approach very similar to Bruner's.

Although the possibilities of controlling the development of cognitive strategies seem definitely promising, the means may not be fully available . . . practically, then if a teacher wants a student to become a good thinker . . . he must provide many opportunities, throughout the course of his instruction, for him to encounter, formulate, and solve problems of many varieties in his chosen field. Such encounters may be expected to lead to the progressive refinement of the cognitive strategies of thinking (p. 522).

The adoption of a problem-solving mode of instruction in this study rests on more than a Gagnéan brand of pragmatism, however. The goal of the present training model is not to teach the medical student the various concepts and principles--of physiology, anatomy, pathology--that are employed in generating problem formulations. Rather it is to increase his capacity to integrate concepts and principles previously learned and to apply these integrated skills to problems of the type encountered in clinical practice. Although much of the research on "discovery versus expository" modes of instruction is equivocal (Shulman and Keislar, 1966), there is at least some research evidence (Egan and Greeno, 1973; Craig, 1969) to suggest that a discovery method may be best suited to achieving cognitive integration (of component skills into problem-solving skills) and lateral transfer (generalization of these new skills to problems not previously encountered). Thus, there is reason to believe that a problem-solving mode of instruction would be the method of choice in the present case, even if expository means were available.

In adopting a problem-solving mode of instruction, one seeks to simulate for the learner problem situations which, in breadth and complexity, closely approximate the real-life situations to which the training is to be generalized. Twelker (1971) has suggested that simulations may be classified into two categories: (1) "interpersonalascendent" (e.g., role playing, simulation games), and (2) "media-ascendent" (e.g., simulator equipment, motion pictures, computer simulations). The present training model involves simulation of the second type: namely, the use of motion picture films to simulate the conditions of the early part of the clinical workup. Each film presents a "physician's eye view" of an encounter with a patient. The full set of films presents eight patients having various medical complaints (within the domain of internal medicine) and diverse demographic characteristics. Films of this type have been used previously as stimulus materials for assessing medical students' clinical skills, but the response mode was a multiple-choice test (Hammond and Kern, 1959). In the present case, the response mode also involves simulation (i.e., the student responds by generating a set of problem formulations as if he were the physician interviewing the patient). Research by Schalock (reviewed in Twelker, 1971) indicates that predictive validity (from a test situation to a complex real-life situation) is greater when simulation is involved in both the stimulus and the response side and,

it is reasonable to assume, that generalization, or, lateral transfer, of skills would also be enhanced by such conditions.

The essence of simulation is a simplification of reality which maintains the aspects of the real-life setting that are necessary to the learning of a task, but omits the aspects that are in some degree unnecessary: in Twelker's (1971, p. 133) words: "Simulation = (real-life) - (taskirrelevant elements)." The use of films to simulate the medical student's encounter with patients relieves the student from having to carry out several activities in which the physician must normally engage: (1) the formulation and asking of questions; and (2) the establishment of rapport and handling of interpersonal relations. Viewing the film does, however, provide the student with the opportunity to carry out the two types of information-processing activity that are of concern in the present study: namely, the detection and encoding of cues on the basis of naturalistic observation (of the patient's physical appearance, manner and speech), and the use of these cues to generate diagnostic problem formulations. In a real-life clinical encounter there is undoubtedly considerable interplay between the formulation and asking of questions and the formulation of diagnostic hypotheses, particularly as the workup progresses and the physician looks for data to test his hypotheses. However, during the first five minutes of

the workup, the physician generally follows a fairly standard questioning format, i.e., elicitation of a description of the patient's current complaint(s); thus, the omission of the task of question-asking should constitute merely a simplification and not a distortion of the conditions under which initial problem formulations are generated.

# The Feedback Component of the Model

Medical problem solving contrasts with many other types of problem solving in that there are no a priori or logical criteria that may be applied in evaluating clinical The criteria that must be used, in evaluating outcomes. final outcomes (i.e., diagnoses) or initial outcomes (i.e., problem formulations), are empirical and a posteriori. For final outcomes there are sometimes specific objective criteria that may be applied, e.g., a laboratory finding that unambiguously substantiates the correctness of a given diagnosis. For initial outcomes, however, appropriateness can only be evaluated in terms of the judgments of experienced practitioners in the field. Thus, the feedback component of the training model is based on data collected from a sample of experienced physicians who carried out the task of generating problem formulations with respect to each of the filmed cases. Data on the physicians' problem formulation outcomes were used to construct "outcome feedback" for each training exercise, and data on the physicians' problem formulation processes to construct "process feedback" for each exercise.

Hammond and his collegues have conducted a number of experiments (Hammond and Summers, 1972; Hammond, Summers and Deane, 1972) in which it was found that the classical type of outcome feedback (i.e., informing the subject of the correct response) was an impediment to improvement of performance on complex tasks (multiple-cue probability learning, clinical judgment) in which the objective is not so much new learning as the application of skills already acquired, or, in Hammond's words, the acquisition of "cognitive control." On the other hand, various forms of "cognitive feedback" were effective in enabling the subject to develop cognitive control. Cognitive feedback is defined as material which "will enable the subjects to perceive not only that their judgment was in error, but <u>why</u> it was in error" (Hammond and Summers, 1972, p. 64).

The outcome feedback utilized in this study includes not only the problem formulations generated by the physician sample, but also a list of the cues that they considered to be relevant to each formulation. The first objective of this feedback is to enable the subject to evaluate the appropriateness of his outcomes (i.e., the formulations he generated as compared to those of the physicians). The second objective of this feedback (and the reason that a cue list is provided for each problem formulation) is to permit the subject to discover some of the reasons his outcomes deviate from those of the physicians. By examining

the feedback material the subject could discover, for example, that he failed to detect certain cues, that he encoded some cues incorrectly, that he failed to "cluster" related cues, that he failed to consider that cues may be relevant to multiple formulations. Thus, the outcome feedback used in this study is much closer to cognitive feedback, as defined by Hammond, than to the classical type of outcome feedback used in learning experiments and programmed instruction. The process feedback utilized in this study is intended to further assist the subject in determining why his outcomes deviate from those of the physicians, and thus is also a form of cognitive feedback. These materials attempt to portray the types of information-processing activity that go on inside the physician's head during the early part of the clinical workup, and, thereby, to enable the subject to discover inadequacies in his own information-processing activities. It is not likely that the medical student will learn to "think like" an experienced physician as a result of receiving process feedback. But, it is possible that he will be able to use information on how physicians think to improve his skill in attaining outcomes similar to those of the physician.

To summarize: The literature reviewed in this chapter suggests that probably the most salient characteristic of medical problem solving is the early generation, and subsequent testing, of diagnostic problem formulations. Although

there is as yet little research evidence regarding the characteristics of the initial problem formulations a physician generates, or the cognitive mechanisms involved in their generation, an attempt was made to define, in general terms, (a) two possible cognitive functions that initial problem formulations may have in the conduct of a clinical workup, and (b) some of the potential risks that could be entailed by the early generation of problem formulations. Finally, consideration was given to issues related to the development of the proposed model for training medical students in the generation of initial problem formulations, including: (1) the choice of a problem-solving (rather than an expository) mode of instruction, (2) the use of films as a means of simulating the conditions of the early part of the clinical encounter, and (3) the use data from experienced physicians to provide the student with cognitive feedback of an outcome and a process nature.

#### CHAPTER III

## METHOD

This chapter describes the methodology of the study. It includes three major sections: (1) production of the films, (2) collection of the physician data, and (3) design of the training experiment conducted with second-year medical students.

## Production of the Films

Eight 16-millimeter films in color and with second were produced. Each film presents the first 4-6 minutes in a physician's encounter with a new patient.<sup>1</sup> The setting of the interview is a doctor's office; thus, the patient is ambulatory, and his problem is of a non-emergency nature.

Since the purpose of the films is to provide the viewer with a realistic simulation of participation in a clinical encounter, they were produced so as to present a "physician's eye view" of the encounter. Throughout the film the camera remains on the patient; the physician's voice is heard but he is never seen. After an initial 30second segment in which the patient walks in and sits down,

<sup>&</sup>lt;sup>1</sup>The average length of the films is 5 minutes, 18 seconds; the films range in length from 4 minutes, 20 seconds to 6 minutes, 10 seconds.

he is shown seated throughout the rest of the film, with close-up shots of his face occurring at several points. In sum, by focusing on the patient, the films were designed to facilitate the viewer's task of adopting the role of physician in a simulated clinical encounter.

The cases for the eight films were selected to represent a cross-section of problems in internal medicine, as well as a variety of patient demographic characteristics (age, sex, occupation). Four of the films are based on cases that were used in the Elstein, <u>et al.</u> (1972) investigation of physician reasoning. The other four cases were developed specifically for this study.

Table 1 lists the eight films, titled according to the demographic characteristics of the patient, and numbered in the order they were presented during the training experiment. The table also indicates the presenting complaint(s) of the patient in each film.

For each film, a case outline was prepared consisting of the following information: (1) information to appear on a written sheet, including the patient's major demographic characteristics (age, sex, occupation) and his temperature; this sheet was presented to subjects viewing the films as having been "filled out by the nurse" just prior to the interview; (2) information to be included in the doctorpatient dialogue, including a list of the patient's complaints, a brief description of the salient attributes of

Number		Title	Presenting Complaint
1	A	21-year-old college senior	Fatigue and weakness
2	A	<b>4</b> 3-year-old landlady of a boarding house	Substernal chest pain
3	A	30-year-old taxi driver	Urinary distress
4	A	40-year-old carpenter	Left chest pain
5	A	<b>19-year-old college</b> sophomore	Headache and sleepiness
6	A	29-year-old lawyer	Low back pain
7	A	57-year-old executive	Cough and fever
8	A	19-year-old student nurse	Abdominal pain and vomiting

each complaint (e.g., onset, duration, location, severity, etc.), and other data of revelance; (3) an indication of significant nonverbal aspects of the case, e.g., the patient's physical appearance, psychological state, etc. An example of a case outline appears in Appendix A.

Production of each film involved the following steps:

1. A case outline was prepared.

2. An experienced amateur actor was selected to play the role of the patient.

3. After the actor had familiarized himself with the case outline, a warm-up session was held in which the actor was coached with respect to his role, i.e., both the verbal presentation of his complaints, and the nonverbal aspects of his role (e.g., gait, posture, gestures, facial expressions, etc.). 4. A trial-run of the interview was videotaped, with immediate replay permitting discussion and critique.

5. The interview was filmed.

Four physicians assisted the author with preparation of the case outlines and coaching of the patient-actors. Two of them played the role of the physician in the films (four films each). Given the constraint that each of the major topics in the case outline be covered, the physicianactor was free to conduct the interview in accordance with his usual practice. He was, however, instructed to utilize a relatively standard, unobtrusive questioning technique, and, in particular, to avoid any questions that would obviously imply a particular problem formulation. During the warm-up session, a general sequence of events was worked out, but, in order to preserve naturalness of dialogue, any tendency to establish a fixed script was avoided.

Each film begins by showing the patient walking into the doctor's office and sitting down to await the arrival of the physician. During this initial segment (30 seconds or less), relevant nonverbal attributes of the patient are presented, e.g., the patient coughs; he is holding his abdomen; he slumps in the chair, etc. When the physician enters, it is assumed that he has with him a sheet filled out by the nurse indicating the patient's name, his major demographic characteristics, and his temperature. The interview begins with the patient

presenting the complaint (or complaints) that have brought him to see the doctor. Questioning by the physician elicits information about the complaint(s), e.g., onset, duration, severity, location, amelioration, etc. As the interview progresses, additional complaints and their attributes, as well as certain other items of relevance, are either presented by the patient or elicited by physician questions. The degree to which information is presented (by the patient) or elicited (by physician questions), as well as the manner in which the patient presents information or responds to questions, vary in each film depending on the type of patient. For example, in Film 1 the patient is a 21-year-old college student who is very articulate in his presentation of information and who responds precisely and in detail to the physician's questions, while in Film 3 the patient is a 30-year-old taxi driver who has a good deal of difficulty in describing his complaints and who responds imprecisely and with minimal detail to the physician's questions.

To summarize: each film presents both verbal and nonverbal information. Information presented in the verbal mode, i.e., the dialogue between the doctor and the patient, consists primarily of a brief review of each of the patient's current complaints (history of present illness), but includes a few items of personal, past medical or family medical history that have particular relevance to the present illness.

The types of nonverbal information presented include: (1) the physical appearance of the patient (e.g., posture, build, dress); (2) the psychological state of the patient, (e.g., gestures, manner of speech, facial expression); and (3) nonverbal cues of particular relevance to the patient's current medical problem (e.g., cough, photophobia, clutching of abdomen).

The objectives of the training experiment influenced the design of the film in several ways. First, each film is intended to provide a brief introduction to a case, on the basis of which the subject is to generate a set of initial problem formulations which he would wish to investigate more thoroughly during the remainder of the workup. The 4-6 minute interview is structured so as to incorporate a limited amount of data on each of the patient's current complaints, but is not intended as a complete history of present illness. This fact is reinforced by the physician's closing statement in the films e.g., "Now, let us go back and review several of your problems." Second, given the fact that the training experiment deals with cognitive information-processing skills, the films are not designed to focus on the affective, interpersonal aspects of the doctor-patient encounter. In producing each film, the physician attempted to use an interview style that represents good medical practice with respect to the establishment of rapport with a patient. The films do not, however, attempt to provide models of interpersonal interaction.

### Collection of Physician Data

The purpose of this phase of the study was to obtain data on physician performance to be used in designing the training experiment involving medical students. The eight films described in the preceding section were shown to a sample of experienced physicians. For each film, two types of data were collected: (1) data on the outcome of the physician's information processing (principally, the set of problem formulations he generated and the cues associated with each), and (2) data on the processes by which the physician generated his set of problem formulations.

## Sample

In utilizing data from a sample of practitioners in a field as the basis for developing materials to train and evaluate students in that field, it would be desirable to obtain a sample of practitioners of proven expertise. In the present study, the ideal physician sample would be a group of clinicians in the field of internal medicine who are known to have outstanding diagnostic skills. Unfortunately, there appears to be no currently available means for identifying expert practitioners in clinical medicine. One method that has been attempted--the use of peer nominations to select "criterial" diagnosticians (Elstein, <u>et al.</u>, 1972)--did not prove to be successful (i.e., the criterial and noncriterial samples were not found to differ on a wide range of problem-solving measures). Thus, in the present

study representativeness rather than criterial expertise was the basis for sample selection.

In selecting the sample the objective was to obtain a group of subjects whose academic backgrounds and clinical experience would be representative of the population of physicians who would generally deal with the type of medical cases presented in the films (i.e., office visits pertaining to problems in internal medicine). It was therefore decided to select a sample of eight physicians that included four specialists in internal medicine (with M.D. degrees) and four family medicine physicians (two with M.D. degrees and two with D.O. degrees). The eight subjects were selected from among the physicians associated with the Michigan State University Colleges of Human and Osteopathic Medicine.

Table 2 summarizes the characteristics of the physician sample, including the type of degrees they hold, their areas of specialization, and the number of years of experience they have had as practicing clinicians, and as medical educators.

Given the lengthy data collection procedure for each film, and the limited amount of time certain subjects could make available, it was not possible to obtain data on the performance of all eight subjects for each of the films. For each film, data were obtained from a minimum of three (out of the four) internists and three (out of the four)
			Average No. Years	Experience
n	Degree	Specialization	As Practicing Clinician	As Medical Educator
4	M.D.	Internal Medicine	11.0	11.7
2	M.D.	Family Practice	8.0	2.5
2	D.O.	Family Practice	8.5	1.5

TABLE 2.--Characteristics of the Physician Sample.

family medicine physicians. For one film (number 6), data were obtained from a total of seven subjects; for the other seven films, data were collected from a total of six subjects.

In order to construct the feedback materials and the dependent variable scoring keys, it was necessary that the physician sample (per film) be of sufficient size (1) to permit identification of commonalities in problem formulation outcomes and processes across the range of academic and clinical backgrounds represented in the sample, and (2) to provide an indication of the range of diversity that would be characteristic of experienced practitioners having such backgrounds. It is believed that the present sample was of sufficient size to provide adequate data on both of these points. The composite set of problem formulations generated for each film always included: (1) a substantial subset of formulations (approximately 30-50% of the composite) which were common to all (or all but one) of the physicians, and thus could provide a basis for determining the outcomes which the student should necessarily attain, and (2) a second superordinate subset of formulations (approximately 80-90% of the composite) which were generated by at least two physicians, and thus could provide a basis for defining the range of outcomes that it would be acceptable for the student to attain.

#### Materials

Two types of materials were used in collecting the physician data: (1) response sheets, and (2) a Process Checklist.

The response sheets were used by the subject to record the problem formulations he had generated while viewing a film. The sheets had a very simple format: a line across the top of the sheet for a problem formulation title, and space underneath for listing the cues of relevance to that title.

In the Elstein, <u>et al.</u> (1972) investigation of physician reasoning, it was found that subjects often had some degree of difficulty in providing an introspective description of the mental processes by which they generated problem formulations. It was therefore decided that, after the subject viewed each film, in addition to tape recording his introspective reconstruction of his thinking process while viewing the film, he would be given a checklist to

fill out. The use of a checklist for the assessment of problem-solving processes was suggested by a study of Marshall (1971).

The Process Checklist devised for this study consisted of a series of 25 statements that pertain to four aspects of the act of generating problem formulations:

- 1. modes of mental representation,
- 2. strategies of problem formulation, including
  - a. initial routines,
  - b. general strategies,
- 3. associative processes of problem formulation,
- 4. cue utilization.

The classification of checklist items according to the above categories, and a brief description of each item appear in Chapter IV (Table 9). A copy of the Process Checklist is contained in Appendix B. In administering the checklist, the subject was instructed to check those items which "characterize your thinking while viewing this film."

The checklist items were derived from data obtained from the physicians who participated in the Elstein, <u>et al.</u> (1972) simulations. The "think aloud" and recall protocols of these physicians were reviewed; statements pertaining to problem formulation processes were picked out, and, with some modification of wording, included in the checklist. In addition, some items were devised to describe processes which were not explicitly stated by the physicians, but which could be inferred from a review of their protocols.

## Procedure

The physician data were collected in individual sessions lasting about three hours during which three or four films were viewed. The session began with the administration of a set of general instructions. For each film, the following steps were involved in the data collection.

Collection of the Outcome Data:

1. The subject was first shown the initial segment of the film in which the patient walks into the doctor's office and sits down to await the arrival of the physician. The film was stopped at this point and the subject was asked to comment on his impression of the patient and on any ideas that came to mind as to what problems the patient might have. The subject's comments were tape recorded.

2. The subject was given the written "nurse's sheet" pertaining to the patient. His comments regarding impressions of the patient or ideas about the patient's problems were tape recorded.

3. The subject was then shown the rest of the film. At the end of the film, he was asked to fill out a response sheet for each problem formulation that he had generated while viewing the film.

4. The subject was asked to provide his tentative assessment of the case. He was asked to indicate: (1) how well substantiated he considered each of his problem formulations to be on the basis of the data obtained;

(2) whether he anticipated that the patient has a single illness or multiple disorders; (3) whether he considered there to be any functional relationships among his problem formulations (e.g., some formulation could be considered to be secondary to, superimposed on, or contributing to, etc., some other formulations). All comments were tape recorded.

Collection of the Process Data:

1. The subject was asked to attempt "to reconstruct your thinking while viewing this film," including such things as the point in the film when each problem formulation came to mind, the cues that were significant in generating each problem formulation, and any revision of initial formulations as the interview progressed. These comments were tape recorded. The physician was offered the opportunity to view the film again as he reconstructed his thinking, but only occasionally did a subject elect to do so.

2. The Process Checklist was administered. The experimenter used the items checked by the subject as a basis for asking additional questions pertaining to processes of problem formulation. The subject's responses were tape recorded.

Although the checklist was administered after each film, it is believed that the length of the list (25 items) and the number of activities intervening between each administration of the list were sufficient to minimize any

effect that exposure to the checklist might have had on subsequent problem formulation activity.

### Analysis

The primary purpose of the analysis of the physician data was to obtain a basis for designing two components of the training experiment: (1) the outcome and process feedback materials for each of the six training films; and (2) the dependent variable scoring keys used to evaluate student performance on the posttest. The development of the feedback materials and scoring keys is described in the next section of this chapter.

Analysis of the physician data was also conducted for a secondary purpose: namely, further specification of the nature of the problem formulation component of medical problem solving, beyond that which has been forthcoming from previous investigations. This analysis was designed to address three questions:

- How early in the clinical workup does the physician begin to generate problem formulations?
- 2. What is the structure of a set of initial problem formulations?
- 3. What cognitive processes are involved in the generation initial problem formulations?

A description of the methods of analysis, as well as a discussion of the results pertaining to each question, are presented in Chapter IV.

# The Training Experiment with Second-Year Medical Students

The training experiment employed a posttest-only control group design, with subjects assigned at random to one of three experimental conditions:

- 1. Treatment I: Training with Outcome Feedback; Posttest
- Treatment II: Training with Outcome and Process Feedback; Posttest
- 3. Control: Posttest only

Since both treatment conditions were based on the general training model described in Chapter I, they had a number of features in common.

 Both conditions consisted of three training sessions (with two films presented at each session), and a posttest session (at which two films were presented).
The general format of each session was the same under both conditions.

2. Under both conditions, the subject carried out the same basic task with respect to each of the films: i.e., having viewed the film, he filled out a set of response sheets indicating the problem formulations he had generated, and he wrote a brief tentative assessment.

3. Under both conditions, the subject was provided with feedback materials based on the physician performance data.

The two training conditions differed, however, with respect to the type of feedback provided. Under one condition (Treatment I), the feedback materials provided the student with "outcome models," i.e., examples of the problem formulations and tentative assessments generated by the physicians for each of the training films. Under the other condition (Treatment II), the feedback materials provided the subject with "outcome models," as defined above, and "process models,"'i.e., materials which portrayed the processes by which the physicians arrived at their problem formulations. The outcome feedback, provided under both treatments, was presented in written booklet form. The process feedback provided under Treatment II included audio supplements to each of the training films, and written materials.

The control condition involved two sessions: (1) an initial orientation session whose purpose will be described subsequently; and (2) a posttest session equivalent to the posttest session under the two training conditions.

#### Experimental Procedure

Under both treatment conditions, training was conducted in three sessions, with two films presented at each session. The order in which the films were presented was the same under both conditions: namely, random order, with the restriction that films involving similar medical complaints and/or similar patient demographic characteristics

were not presented consecutively (see Table 1). Under both treatment conditions, the instructional sequence followed for each film remained the same across the three training sessions. The posttest session, involved the presentation of two films, with the same procedure followed under all three experimental conditions.

All experimental manipulations were administered to the subjects by means of individual booklets in selfinstructional format. At the beginning of the first training session, this booklet provided the subject with a set of orientation materials designed to acquaint him with the problem formulation task. At the beginning of sessions two, three and four, the subject was given review materials. The instructional sequence (or posttest task) for each of two films was then administered by means of the self-instructional booklet. Thus, the role of the experimenter was limited to a small number of preliminary verbal instruction.

It was decided that a single session for the subjects under the control condition would be undesirable for two reasons. (1) Because a single control group session would require administration of the orientation materials prior to the presentation of the two posttest films, it would be longer than the treatment group posttest session, and, of course, not fully equivalent in content. (2) It is possible that on a subject's first exposure to the task his performance might be depressed, or affected in some other

unknown way, by the novelty of viewing a filmed interview and then having to fill out a set of relatively unfamiliar response sheets. It was believed that any "novelty effect" would be eliminated after a subject's first exposure to one of the films, and to the task of filling out the response In order to make the posttest session conditions sheets. as similar as possible across all three groups, and in order to control for the possibility of a "novelty effect," it was decided to conduct two sessions for control group subjects. During the first session, the subject was provided with orientation materials (similar to those administered at the first session under the treatment conditions), and he carried out a task designed to control for a possible novelty effect during the subsequent posttest session: namely, the sixth training film was presented, and the subject recorded his problem formulations and tentative (Feedback was, of course, not provided.) assessment. The second control group session involved the same procedure as the treatment group posttest sessions: review of the orientation materials, followed by presentation of the two The format of the experimental procedure posttest films. is outlined in Table 3.

The heavy academic work-load of the second-year medical student population placed several constraints on the scheduling of the experimental sessions. Because of the students' very full class schedule during the day, it

,

	Experimental Conditions			
Week	Treatments I and II	Control		
1	Training sessionl Orientation Film l Film 2			
2	Training session2 Review Film 3 Film 4			
3	Training session3 Review Film 5 Film 6	Orientation session Orientation Film 6		
4	Posttest session Review Film 7 Film 8	Posttest session Review Film 7 Film 8		

TABLE 3.--The Experimental Procedure.

was necessary to conduct the sessions in the evenings. Moreover, because of the limited number of time slots available for scheduling sessions (i.e., four weekday evenings), it was necessary to conduct group sessions involving all 16 of the subjects assigned to an experimental condition. Although individual administration of the training and post test was not feasible due to practical constraints, it was not considered to be necessary (in order to treat the subject as the unit of analysis) since all experimental manipulations were carried out by means of a self-instructional booklet, and there was no interaction among subjects during the sessions.

The three experimental conditions were randomly assigned to three evening time slots. The treatment group sessions were held on four consecutive weeks. The control group sessions were held on two consecutive weeks (during the third and fourth weeks of the treatment group sessions). Although this procedure had the undesirable feature of confounding each experimental condition with one weekly time slot, it was not considered that this factor would pose a serious threat to the internal validity of the experiment. Under the treatment conditions, the first session lasted approximately two and one-half hours, and the other sessions approximately two hours each. Under the control condition, the first session lasted one and one-half hours, and the second two hours. A 10-minute break was taken about half-way through each session. An assistant aided the experimenter with the administration of each session.

Due to absences at the scheduled group sessions, it was necessary to conduct make-up sessions on an individual or small-group basis. A total of 10 subjects (4 from Treatment I, 3 from Treatment II, and 3 from the control condition) participated in one make-up session; one Treatment I subject participated in two such sessions. Each make-up session was held within two days of the scheduled group session it was replacing.

### The Problem Formulation Task

For each of the filmed training and posttest cases, the subject was confronted with the same basic task. Before viewing the film, he was given the following instructions:

While viewing the film, you should generate a set of initial problem formulations which you would want to investigate more thoroughly if you were to continue the workup beyond the first 4-6 minutes presented in the film.

After the subject read the "nurse's sheet" and viewed the film, he recorded the problem formulations he had generated on response sheets. The response sheets were set up so that a group of hierarchically related problem formulations would be recorded on a single response sheet, with the most general formulation in the hierarchy listed on the front of the response sheet, and the more specific formulations listed on the back. If a subject generated a formulation that was not hierarchically related to another formulation, it was listed on the front of the response sheet. An example of a response sheet appears on page 65.<sup>1</sup>

<sup>&</sup>lt;sup>1</sup>In the training materials, the term "problem formulation" was used to designate the response recorded on the front of the response sheet, and the term "more specific diagnostic possibility" to designate responses (if any) recorded on the back of the sheet. This distinction was made in order that the terminology used in training would be consistent with that to which the students had become accustomed in learning to fill out a Problemoriented Record. However, unless specifically indicated, a single term--"problem formulation"--has been employed throughout this report to refer to any diagnostic hypothesis, whether listed on the front or back of a response sheet.

Film		
 Problem formulation title:		
CUE LIST		
More specific diagnostic possibilities (if any) are to be listed on the back of the sheet.		

.

~

	Title	Cues of particular relevance
1.		
2.		
		·
3.		
4.		
5.		
6.		

More specific diagnostic possibilities which you have considered for this problem formulation (if any)

.

After the subject had filled out the problem formulation response sheets, he wrote a brief paragraph giving his tentative assessment of the case. He was instructed that his tentative assessment should indicate:

- --how well substantiated you consider each of your problem formulations to be on the basis of the data obtained thus far;
- --whether you anticipate that the patient has a single illness that will account for his various problems, or that he has multiple disorders;
- --whether you consider there to be any relationships among your problem formulations. For example, you may consider one problem formulation to be secondary to, superimposed on, or contributing to, etc. some other formulation.

### The Instructional Sequence

Under both treatment conditions, the same instructional sequence was followed for each of the six filmed training cases. This sequence involved five steps which are summarized below.

- STEP 1: The subject read the "nurse's sheet" for the patient in the film.
- STEP 2: The subjects viewed the film of the 4-6 minute interview with the patient.
- STEP 3: The subject recorded the problem formulations he had generated, and wrote a brief tentative assessment.
- STEP 4: The subject was provided with feedback on the performance of the experienced physicians.

a. "Feedback Sheet 1" was presented.

b. Treatment I:

The film of the interview was presented a second time. Treatment II: The process feedback version of the film was presented.

c. "Feedback Sheet 2" was presented.

STEP 5: The subject filled out a self-evaluation checklist.

The first three steps in the sequence constituted the basic experimental task. Step 4, which embodied the experimental manipulation of feedback, was the only step in the instructional sequence which differed for the two treatment conditions.

The feedback was presented to the subjects in three parts. The first part (Feedback Sheet 1, entitled "Major Problem Formulations") was designed to provide the subject with feedback on the problem formulation <u>outcomes</u> that were <u>common</u> to the responses of all, or nearly all, of the physicians. This sheet enabled the subject to determine whether he had generated those formulations which the physician data indicated were of major importance for the case under consideration.

The second part of the feedback consisted of a film presentation: in the case of Treatment I, a re-presentation of the standard film of the interview; in the case of

Treatment II, a presentation of the process feedback version of the film. Under both treatment conditions, a second exposure to the interview provided the subject with implicit feedback on the adequacy of his detection and recall of cues. The conditions differed, however, with respect to the provision of process feedback. The film presented under the Treatment II condition provided explicit feedback, via the physician's "think aloud" comments, on the processes by which the physician sample generated each of the problem formulations listed on Feedback Sheet 1. Under the Treatment I condition, on the other hand, the second presentation of the standard film provided the student with the opportunity to attempt to reconstruct, on his own, the processes by which the problem formulations on Feedback Sheet 1 were generated.

The third part of the feedback (Feedback Sheet 2) contained one section that was the same for subjects under both conditions. This section (entitled "Additional Problem Formulations") was designed to provide the subject with feedback on the <u>range</u> of diversity in the physicians' problem formulation <u>outcomes</u>.

The second section of Feedback Sheet 2 (entitled "Summary") differed for the two treatment groups. Under Treatment I, it summarized the comments included in the physicians' tentative assessments. Under Treatment II, it consisted of a reconstruction of the physicians' reasoning

about the case, including a description of the processes by which the problem formulations listed on both feedback sheets were generated, as well as a summary of the comments in the physicians' tentative assessments. Feedback Sheet 2 pointed out both the commonalities and the range of diversity that were characteristic of the physicians' problem formulation processes, and their tentative assessments.

Table 4 outlines the properties of the feedback presented under the two treatment conditions.

freatment II
Outcomes (C)
plemented Film: Processes (C)
Outcomes (D)
Processes (D) and Outcomes (C,D)

TABLE 4.--Properties of the Feedback Presented under the two Treatment Conditions.

NOTE: PF = problem formulation; TA = tentative assessment; C = feedback indicating the commonalities found in the performance of all or nearly all physicians; D = feedback indicating the range of diversity found in the performance of the physicians.

The fifth and final step in the instructional sequence consisted of filling out a self-evaluation checklist. This checklist was designed to serve two functions: (1) to insure that the subject carried out the process of comparing his own performance to that of the experienced physicians, and (2) to provide the subject with a sense of closure at the completion of the instructional sequence for a case. The first part of the checklist listed the titles of the problem formulations generated by the physicians. The second part of the checklist listed the statements regarding functional relationships between problem formulations that were found in the physicians' tentative assessments. The subject was instructed to check each item in the list that corresponded to one of his own responses.

### Outcome Feedback

These feedback materials were designed to provide the subject with models of appropriate outcomes for each of the filmed training cases: namely, the problem formulations and tentative assessments generated by the experienced physicians.

Feedback Sheet 1 and the first section of Feedback Sheet 2 were based on a tabulation of the physician problem formulation data. Feedback Sheet 1 presented those problem formulations generated by at least five of the physicians who viewed the film. The first section of Feedback Sheet 2 presented the problem formulations generated by two to four of the physicians. Responses generated by only one physician were not included in the feedback.

Each problem formulation presented in the feedback sheet included two components: (1) a problem formulation title, and (2) a list of cues of relevance to that title. Problem formulations were organized on the feedback sheets so as to indicate hierarchical relationships (if any) among subsets of formulations. Cues of relevance to all formulations in a hierarchy were listed under the problem formulation title at the head of the hierarchy; only those cues of particular relevance to each more specific formulation in the hierarchy were listed under the subordinate title(s) in the hierarchy (see Figure 1).

The second section of Feedback Sheet 2 was based on a review of the transcriptions of the physicians' taperecorded tentative assessment comments. It indicated both the commonalities and the range diversity in the physicians' assessments with respect to the three topics listed on p. 67.

#### Process Feedback

These feedback materials were designed to provide the subject with models of the processes by which the physicians arrived at their problem formulation outcomes.

For each case, a process feedback version of the training film was produced. Each of these films included three or four "think aloud" segments interposed at appropriate points in the standard film of the interview. For each such segment the dialogue between the doctor and

GI DISORDER: ULCERATIVE COLITIS, or REGIONAL ENTERITIS/ILEITIS

GI disorder pain in right lower quadrant of abdomen, for 4 months occurs in evening, lasting several hours not relieved by aspirin or Darvon not related to foods diarrhea: increase in number of stools, from 1 to 4-5/day, over 3 month period mucous in stools blood in stools pieces of food in stools weight loss, 25 lbs. in 1-2 months good appetite, eating more than usual extreme fatigue and weakness, for 2 months no vomiting Ulcerative colitis, or regional enteritis/ileitis diarrhea blood and mucous in stools weight loss of 25 lbs., with good appetite age 21 college senior: under academic stress concerned about keeping up with studies

Figure 1. A Sample of the Outcome Feedback for Film 1.

patient stopped, an image of the patient was frozen on the screen, and the physician was heard "thinking aloud" about such matters as:

- --the problem formulations he had generated up to that point;
- --the cues which had led to the generation of these formulations;
- --alternative interpretations being considered for certain ambiguous cues;

--any strategies that were guiding his thinking; --his impressions of the patient;

--his revisions of previous formulations in the light of new data.

In sum, the "think aloud" segments attempted to provide the viewer with simulated access to the processes going on in the physician's head as he conducted the interview.

The process feedback films were produced as follows. For each film, the physicians' retrospective recall protocols and their responses to the Process Checklist were summarized in outline form. On the basis of these data, three or four appropriate stopping points in the film dialogue were selected, and a script was prepared for the "think aloud" segment to take place at each point. Since the process feedback films were intended to portray the cognitive activities characteristic of experienced physicians in general, the scripts for the "think aloud" segments included only those processes typically involved in the generation of the "major" problem formulations listed on Feedback Sheet 1.

The recording of the "think aloud" segments was carried out under conditions similar to those of the filming of the interviews. (1) In order to preserve naturalness of speech in the "think aloud" segments, the physician was provided with a script outlining the topics to be covered in each segment, but was otherwise free to improvise the comments he would make if he were in fact "thinking aloud."

(2) After the physician had reviewed the script, a trial run was carried out. (3) The final tape recording of the "think aloud" segments was undertaken. In order to facilitate the physician's task in simulating spontaneous "thinking aloud," he was shown the film of the interview up to each stopping point just prior to carrying out each commentary. Finally, a film laboratory undertook the production of the freeze frames of the patient and the insertion of the "think aloud" auditory tracts into the original films of the doctor-patient interview.

Further process feedback was provided to the subject in written form by means of the "Summary" on Feedback Sheet 2. This summary attempted to indicate both the commonalities and the range of diversity of the physicians' problem formulation processes. It included: (1) a recapitulation of the main points included in the "think aloud" segments of the process feedback film (relevant to the "major" problem formulations on Feedback Sheet 1), and (2) a presentation of the processes underlying the generation of the "additional" problem formulations (found on Feedback Sheet 2). The "Summary" also included discussion of the physicians' tentative assessments. Thus, the final portion of the feedback materials provided to subjects under the Treatment II condition attempted to integrate all feedback information of relevance to the case.

# Materials

All instructions and written training materials were provided to the subject by means of a booklet in selfinstructional format. The first section of the booklet (entitled "Introduction") provided the subject with the orientation materials. For the treatment groups, these materials included: (1) a statement of the purpose of the instructional package; (2) a description of the role of initial problem formulations in medical problem solving; (3) a definition of the components of an initial problem formulation; (4) a definition of a tentative assessment; (5) a description of the materials to be used; (6) a summary of the five steps in the instructional sequence to be followed for each filmed training case; (7) a set of guidelines to follow in filling out the problem formulation response sheets; (8) examples of two problem formulation sheets and a tentative assessment sheet filled out for a sample case. The control group received the same orientation materials, except that item 6 was omitted and the other items were modified as needed to omit reference to the feedback materials.

The section of the booklet for each of the six training films had an identical format: instructions were presented for each of the five steps in the instructional sequence, and the written training materials (nurse's sheets, and feedback sheets) were presented at the appropriate points in the sequence.

The sections of the booklet for the two posttest films (and for the control group orientation film) presented instructions equivalent to those for steps one, two and three in the instructional sequence.

At the beginning of each new session, the subject received guidelines for review of the materials presented at the preceding session(s). At the beginning of session four, the experimental subjects also received a summary feedback sheet entitled "Common errors observed in your responses to the previous cases." This sheet was prepared by the experimenter after having reviewed all subjects' responses to the six training films. The sheet discussed the following types of errors:

- errors in the generation of problem formulation titles due to failures to organize information in an appropriate manner (items 1 and 2).
- errors in the listing of cues under problem formulation titles due to insufficient consideration of cue relevance (items 3, 4 and 5).
- errors in writing a tentative assessment due to insufficient consideration of relationships among problem formulations (item 6).

Along with a description of each type of error, a guideline for avoiding the error was presented.

In addition to the Instructional Booklet, the subject received a Response Booklet. This booklet contained

the response sheets to be used in recording the problem forjulations and tentative assessment generated for each case. It also included the self-evaluation checklists for each of the six training cases.

Examples of the experimental materials are contained in Appendix C.

### Posttest Tasks

The Basic Posttest Task.--At the posttest session the subject carried out the basic problem formulation task with respect to each of two films. Selection of the films to be used for the posttest was based on two considerations. First, in order to obtain as broad a sample as possible of the content domain (i.e., internal medicine), films that presented dissimilar cases, with respect to type of medical complaints and patient demographic characteristics, were selected. Second, in order to avoid a possible ceiling effect, the films were selected from among those which were found, upon examination of the physician data, to provide the basis for the generation of a relatively large number of problem formulations.

The Additional Posttest Tasks. After the subject had completed the basic posttest tasks, he was administered two additional tasks pertaining to the second posttest film. The purpose of these tasks was to determine the extent to which perceptual and memory factors (i.e., factors in the processes of detecting, encoding and retrieving cues

rather than in the process of generating problem formulations per se) may have affected the subject's performance on the basic posttest task. Adequate detection, encoding and retrieval of cues is obviously a necessary prerequisite for generating problem formulations. A high level of performance on the basic posttest tasks would imply that these prerequisites were met. However, a low level of performance would be open to three interpretations: (1) failure in the process of generating problem formulations; (2) failure in the perceptual and memory processes that are prerequisites for generating problem formulations; (3) failure in both domains. In order to more precisely assess between-group differences on the basic posttest task, two additional tasks were devised. Both tasks pertained to the second Although it would have been of interest posttest film. to administer these tasks for the first posttest film as well, this possibility was rejected because of the risk that interpolation of the additional tasks might influence subjects' performance on the second basic posttest task.

The first additional task (Recognition of Cues) required the subject to indicate on a checklist those cues he recalled being presented in film 8. The checklist contained 64 randomly ordered items, 32 of which were cues presented in the film, and 32 of which were distractors. The cues included in the checklist were items which had been listed by at least one member of the physician sample

as relevant to some problem formulation he generated while viewing the film. The distractors were devised by the experimenter in collaboration with a physician consultant, and were designed to represent three plausible but incorrect types of data.

1. consistent distractors: 16 items which were not presented in the film but which would be consistent with the cues that were presented.

2. contradictory distractors: 8 items which contradicted a cue that was presented in the film.

3. inconsistent distractors: 8 items which were not presented in the film and which would be inconsistent with the cues that were presented.

After completion of the above task, the subject carried out a second additional task (Additions to Response Sheets). For this task he was provided with a list of the 32 cues that had been presented in film 8. After reading the list, he was instructed to make any additions he believed appropriate to his problem formulation response sheets, including: (1) addition of cues to the problem formulations he had previously recorded; and (2) addition of new problem formulations which he thought of after reading the cue list.<sup>1</sup>

<sup>&</sup>lt;sup>1</sup>These additions were made in ink so as to be readily distinguished from the subject's initial responses in pencil.

In sum, the first additional task was designed to determine whether failure to detect, encode and recall cues placed constraints on the subject's performance of the basic posttest task, while the second additional task was designed to ascertain whether the removal of potential perceptual-memory constraints would permit the subject to improve his problem formulation performance. Copies of the materials for the two additional posttest tasks are found in Appendix D.

Several comments are in order regarding the interpretation of performance on the additional tasks. Subjects were allowed to take notes while viewing the films, and nearly all did so. Thus, it is primarily failures in the detection and/or encoding of cues, and only secondarily failures in the retrieval process, that are at issue here. Secondly, performance on the second additional task does not, in itself, provide a pure measure of what the subject's performance would have been in the absence of perceptualmemory constraints. A subject may be able to generate additional problem formulations on this task not because he failed to detect and/or encode relevant cues in viewing the film, but simply because the additional task provides a second exposure to the cues. Thus, in interpreting performance on the second additional task, recourse must be made to other sources of data (e.g., items checked on the recognition task) in order to infer that it was failure

to detect and encode cues that inhibited a subject's initial problem formulation performance.

The Questionnaire. -- (Treatment groups only) After the subject had carried out the two additional tasks described above, he was given a questionnaire to fill out. The questionnaire contained four sections. Section one included statements to which the subject responsed on a five-point scale (strongly agree, agree, no opinion, disagree, strongly disagree). These statements were designed to elicit the subject's opinion of, or attitude toward, various aspects of the training procedure: including the instructional booklet, the films of the interviews, and the feedback materials. The questionnaire also sought the subject's opinion concerning the appropriateness of the materials for second-year students, the degree to which the materials had been effective in improving his problem formulation skills, and several possible ways of integrating the materials into the current curriculum. The second section consisted of a checklist designed to determine the degree to which the subject pursued an interest in the training cases outside of the experimental sessions, either through discussion with students and/or faculty, or by looking up reference materials. The third section of the questionnaire was an open-ended request for comments and suggestions. The fourth section requested for subject to report any clinical experience he had had, prior to

participating in the experiment, that involved contact with patients. He was asked to indicate both the type of experience (e.g., as an intern, physician's assistant, medic, nurse) and the extent of the experience (hours/week, and number of weeks). These latter data were collected for the purpose of sample description, and to ascertain the degree to which prior experience may have affected the experimental outcomes.

A copy of the questionnaire is contained in Appendix E.

## Subjects

A sample of 48 students in the third term of their second year of medical school participated in the experiment. The decision to sample second-year students was based on consideration of three criteria. First, it was believed that training in the generation of problem formulations utilizing filmed case presentations would be most appropriate prior to the student's participation in clinical clerkships (i.e., before summer term of his second year). Second, in order for the training to be effective it was necessary that the student have acquired sufficient medical science background to be able to deal with the range of cases in internal medicine presented by the films. Third, it was necessary that the student had not already attained a level of skill in the experimental task that would

preclude improvement in this skill via participation in the training experiment.

Since the experiment was to be conducted spring quarter, there were two potential populations which met the first sampling criterion: students enrolled in the third term of either their first or second year in the College of Human Medicine, Michigan State University. It was believed that the first-year student would not have mastered sufficient medical content to meet the second sampling cri-Second-year students, on the other hand, clearly terion. met this criterion. During the two and one-half terms of their Focal Problems course they had dealt with written case materials of relevance to each of the eight experimental films.<sup>1</sup> Thus, a primary concern became to determine whether the second-year student population would meet the third sampling criterion. The results of the pilot testing indicated that there did not appear to be a substantial risk of a ceiling effect due to the second-year students' prior problem-solving experience (see Table 6). Thus, second-year medical students were chosen as the target population for the experiment.

<sup>&</sup>lt;sup>1</sup>The following "focal problems" involved medical content of relevance to the experimental films: polyuria (films 2,3,8), headache (film 5), painful joints (film 6), diarrhea (film 1), shortness of breath (film 6), fever (films 3,5,7), anemia (films 1,4), chest pain (films 2,4), hematuria (film 3), cough (film 7), abdominal pain (films 1,3,8).

The list of students enrolled in the third term of their second year in the College of Human Medicine included 81 persons. Four students who had begun their primary clerkship that term were eliminated from the list prior to sampling; thus, the target population included 77 persons.<sup>1</sup> Of these, five persons participated in the pilot testing, resulting in a population of 72 students from which the experimental sample was selected. Students were randomly selected from the list, and randomly assigned to one of the three experimental conditions. Each person selected was contacted by telephone to determine whether he could participate in the experiment on the dates scheduled for the experimental condition to which he had been assigned. In four instances, a subject was willing to participate but not available on the dates in question. He was therefore randomly assigned to one of the two other conditions. In order to obtain a sample of 48 participants, a total of 68 students were sampled. The students sampled who were unable to participate included 17 refusals, and 3 students who could not be contacted by telephone. Thus, a participation rate of 70.6% was obtained. On contacting each person he was told that he would be paid approximately \$5.00/hour

<sup>&</sup>lt;sup>1</sup>It was found in reviewing the subjects' responses to section four of the questionnaire that one subject who had begun his primary clerkship was inadvertantly included in the sample (Treatment I condition). Since this subject's scores on the dependent variables did not substantially differ from those of the rest of the sample, he was not excluded from the analysis.
for participation in the experiment. It is believed that without payment the refusal rate would have been considerably higher, and attrition would have occurred across the training sessions.

Ten of the students who refused to participate in the experiment gave the reason of being "too busy" with their studies. The remaining seven students gave a variety of reasons, including prior committments or lack of interest. In order to determine whether those who refused to participate (or who could not be contacted) differed systematically from the sample of participants, it was possible to examine the students' final exam scores for the Focal Problems course in which they were enrolled the term preceding the experiment. These scores were selected for examination since they would provide the best available measure of the students' pre-experimental level of achievement in areas that could be expected to correlate substantially with the experimental posttest variables.<sup>1</sup> Inspection of these data **r**evealed that the four students with exam scores that were **considerably lower than all other students (i.e., scores** in the 40's, whereas the range of all other students' Scores was 61-87) were among the refusals. Thus, it would appear that the weakest students in the population may have systematically eliminated themselves from participation

<sup>&</sup>lt;sup>1</sup>The Focal Problems exam is described in more detail on p. 92.

in the experiment. However, with these four students excluded, the mean exam scores for the participant and refusal/no contact groups were virtually identical (72.06 and 72.67, respectively). A 90% confidence interval calculated on the difference between these means indicated: (1) that zero was contained well within the bounds of the interval, and (2) that the range of values spanned by the interval was very narrow (see Table 28, Appendix H). This finding lends support to the argument that, with the exception of the students in the bottom 6% of the class, the results of the experiment may be generalized to the target population of second-year medical students.

The target population of real interest, however, extends beyond the 77 second-year medical students enrolled at Michigan State. One would wish, utilizing the arguments of Cornfield and Tukey (1956), to generalize the results of the experiment to a hypothetical target population that includes all second-year medical students that are similar to those actually sampled. In so doing, it is of course necessary to take into consideration the characteristics of the sample at both the individual and the institutional level. Table 5 presents selected characteristics of the individuals included in the sample. Since amount of prior clinical experience (e.g., as an extern, physician's assistant, nurse, medic) may affect the degree to which students are able to profit from the simulated clinical

exercises, this variable is reported, along with sex, age and Medical College Admission Test (MCAT) scores.

At the institutional level it is necessary to consider the ways in which the first two years of the curriculum at Michigan State may differ from that found elsewhere. In addition to courses in the basic sciences which are probably quite similar to those at other institutions, the Michigan State students take, during their first two years, several courses which focus on the development of ' clinical skills:

 The Doctor-Patient Relationship course (1 quarter, 3 hours per week), which emphasizes the development of interpersonal skills;

2. The Clinical Sciences course (3 quarters, 1 afternoon per week), which involves supervised contact with patients, including history taking and physical examination.

3. The Focal Problems course (3 quarters, 2-3 hours per week), which focuses on a problem-oriented approach to clinical medicine, including the development of problemsolving skills and use of the Problem-oriented Record. Thus, all participants in the experiment had a substantial degree of prior coursework having a clinical orientation. In their Focal Problems course, in particular, they had become familiar with the concepts of utilizing cues elicited during a clinical workup in order to generate a list of problem

Sex	Male: $n = 42$	Female: $n = 6$
Age	Mean: 25.5	Standard deviation: 2.5
MCAT Scores	Mean:	Standard deviation:
Verbal Ability Quantitative	523.1	96.5
Ability General	552.6	80.8
Information Science	545.5 537.9	72.2 84.3
Prior Clinical Experience (40 hour weeks)	None: n = 8 1-12 wks.: n = 15 13-52 wks.: n = 3 52 or more wks.: n = 6	

TABLE 5.--Selected Characteristics of the Student Sample.<sup>a</sup>

<sup>a</sup>MCAT scores were available for only 42 subjects. Data on prior clinical experience were obtained from the questionnaire administered to the treatment groups (n - 32).

formulations as entries in a Problem-oriented Record. However, their experience in this course differed from that provided by the training experiment in several ways: (1) the course exercises were based on written case summaries; (2) the course exercises did not focus on the generation of initial problem formulations based on cues obtained during the earliest part of the workup: (3) feedback consisted of the course instructor's evaluation of their performance.

# Pilot Testing

The experimental procedure and materials were pilot tested with five subjects randomly selected from the

population of second-year medical students who were to participate in the experiment. An initial version of the Treatment I materials for two films were administered to two subjects. Subsequent to this initial pilot test, all experimental materials were prepared, and a second pilot test was conducted with four subjects (including one person who participated in the initial pilot test, plus three additional subjects). Two subjects were administered the orientation materials and the materials for two films under the Treatment I condition, and two subjects were administered the posttest materials.

The pilot testing served two purposes. First, on the basis of the subjects' comments and criticisms, certain revisions were made in the materials. Secondly, on the basis of the subjects' performance, it was ascertained that the training would be appropriate for a second-year medical student population. The subjects' Problem Formulation scores (reported in Table 6) indicated that they easily met the second sampling criterion (i.e., prior mastery of prerequisite medical content), and that on the posttest films in particular they adequately met the third sampling criterion (i.e., a level of skill in carrying out the experimental task that was low enough to permit detection of a training effect).

Subject	Film	PF Score <sup>a</sup>	Maximum Possible Score
1	2	38	6 0
	4	30	6 0
2	2	34	6 0
	4	36	6 0
3	7	14	54
	8	12	78
4	7	22	54
	8	31	78

TABLE 6.--Results of the Pilot Test.

<sup>a</sup>The PF (problem formulation) score is defined on pp. 96-98.

# The Covariate

In a number of studies of medical problem solving (e.g., Elstein, <u>et al.</u>, 1973; Gordon, 1973), a high degree of variability on the dependent measures has been found. Thus, it was considered quite important that measures on an appropriate covariable be obtained in order to increase the precision of the statistical analysis.

Probably the best measure would have been a pretest in which the subject carried out the same task as on the posttest. However, this possibility was rejected for the following reasons. First, it was believed that in order to obtain reliable pretest measures on the dependent variables at least two filmed cases would have to be employed. This would have reduced the number of films available for training from six to four. It was believed insufficient training would be even more likely than lack of precision to result in a nonsignificant treatment effect. Moreover, a nonsignificant outcome due to failure to carry out an adequate test of the treatment would be a more serious experimental failure than the occurrence of a Type II error due to lack of precision. It was therefore decided to attempt to obtain covariate measures from some other source than a pretest.

The source decided upon was the final exam in the Focal Problems course in which all second-year medical students had been enrolled the term prior to the administration of the experiment. This measure was selected for several reasons. First, since it was a 100-item long exam of multiple-choice and true-false questions, it could be expected to provide relatively reliable and objective data. Second, more than any other available measure, it could be expected to correlate with the dependent variables in the present study. Although the majority of items in the test were designed to measure the student's knowledge of medical science content that had been covered in the course, a substantial number of items attempted to assess his ability to use data regarding a patient to make a differential diagnosis, or to select diagnostic or theraputic options. Thus, it was anticipated that use of this exam as a covariate would be effective in reducing within-group

variability primarily on the dimension "pre-experimental knowledge of medical science content" and secondarily on the dimension "pre-experimental ability in solving medical problems."

### Dependent Variables

A subject's performance on the basic posttest tasks was evaluated in terms of four dependent measures:

- 1. a problem formulation score (PF)
- 2. a cue utilization score (CUE)
- a classification of cues with respect to problem formulations score (CUE-PF)
- 4. relationships among problem formulations (R-PF)

For each variable, the adequacy of the subject's performance was measured by means of a scoring key derived from the physician performance data. The subject's score on each variable was calculated separately for each of the two posttest tasks, but, for purposes of statistical analysis, his scores were summed across tasks, thus yielding four dependent measures per subject.

Each scoring key was designed to measure the degree to which the student's performance on a given variable approximated that of the experienced physicians. Each key contained a list of various potential responses, with points assigned to each. The number of points assigned to a response was weighted to reflect the relative frequency with which the response occurred in the experienced physician data. Certain additions to the keys, as well as validation of certain components in the keys, were carried out by means of independent consultations with two additional physicians who were not part of the sample of eight.

Three of the dependent variables (PF, CUE, CUE-PF) were based on the information the subject recorded on his problem formulation response sheets. Each of these variables pertained to one component of the interrelated cognitive outcomes which resulted from the subject's simulated encounter with a patient. As depicted in Figure 2: the CUE score pertained to the functional data base (i.e., set of cues) which the subject extracted from the film and utilized to generate problem formulations; the PF score pertained to the set of problem formulations he generated; the CUE-PF pertained to the way in which he classified the cues he obtained with respect to the problem formulations he generated.

The fourth dependent variable (R-PF) was based on information the subject recorded in his tentative assessment, and pertained to functional relationships which he hypothesized to exist between problem formulations he had generated.

The remainder of this section will be devoted to a discussion of the properties of each score and of the general principles underlying the construction of the scoring keys. For more detail the reader is referred to the copies of each key and of the scoring instructions, found in Appendix F.



Figure 2. Relationships between Cognitive Outcomes and the Dependent Variable Scores CUE, PF and CUE-PF.

<u>CUE score</u>.--This score was designed to measure the adequacy of the functional data base which the subject extracted from the film and utilized in the generation of problem formulations. It was based on the cues which the subject listed on his response sheets, irrespective of the problem formulation title(s) under which he listed them. The key consisted: (1) of a list of all cues that were utilized (i.e., listed under any problem formulation title) by the physicians, and (2) for each cue the number of points to be obtained if the subject utilized the cue. Points were allotted to cues as follows:

No. of physicians utilizing the cue	Points
n = 5-6	3
n = 2-4	2
n = 1	1

A subject's CUE score consisted of the sum of points obtained for each cue he utilized. The CUE scoring key for film 7 included 20 items, and yielded a maximum possible score of 42. The CUE key for film 8 included 32 items and yielded a maximum possible score of 61. Thus, the range of scores on the CUE variable (summed over both posttest tasks) was 0-103.

<u>PF score</u>.--This score was designed to measure the appropriateness and thoroughness of the subject's set of problem formulations. The scoring key included a list of all problem formulation titles generated by the physician sample, and, in addition, a list of "other" titles (i.e., student responses not occurring in the physician data) that were judged to be acceptable by both of the physician consultants. Points were allotted to each title as follows:

No. who	of physicians listed the title	Points
n =	5-6	6
n =	: 2-4	3
n =	<ul> <li>l (or, judged acceptable by the two physician consultants)</li> </ul>	1

A subjects' PF score consisted of the sum of points obtained for each problem formulation title he recorded.<sup>1</sup> The PF key for film 7 included 13 titles (plus a list of 12 "Other Acceptable Responses"), and yielded a maximum possible score of 54 points. The PF key for film 8 included 20 titles (plus a list of 18 "Other Acceptable Responses"), and yielded a maximum possible score of 78 points. Thus, the range of scores of the PF variable (summed over posttest tasks) was 0-132.

The performance dimension of thoroughness was taken into consideration both in the construction of the key and in the definition of the PF score as cumulative summation. The scoring key was constructed to incorporate as wide a range of potential formulations as possible: all of the physicians' responses were included, and "other acceptable responses" (as judged by the physician consultants) were added so that restrictions due to sampling error in the original physician data would be minimized. By cummulatively summing points across the problem formulation titles a subject generated, the thoroughness of his performance (i.e., the number of problem formulations generated) would be reflected in the resultant score.

<sup>&</sup>lt;sup>1</sup>The term "problem formulation title" refers to both the title listed on the front of each problem formulation response sheet and the titles of more specific diagnostic possibilities (if any) listed on the back of each response sheet.

A second performance dimension--appropriatness-was incorporated in the key in several ways. First. the points assigned to each title in the key were weighted to reflect their relative frequency of occurrence in the physician sample. Second, a set of codes was included in the key indicating the permissible ways in which hierarchically related titles could be recorded. If a subject recorded a title in a way that was not permissible (e.g., he recorded "pheumonia" as a diagnostic possibility on the back of a problem formulation sheet with the title "cancer"), he received no points for this title. Third, in order to control for inflation of the PF score due to a tendency on the part of a subject to "catalogue" every diagnostic possibility he could think of, two features were included in the key: (1) a title was not scored if there were no cues listed with it; (2) the number of points that could be obtained for "other acceptable responses" was restricted to a maximum of six (i.e., six one-point responses).

<u>CUE-PF score</u>.--The CUE-PF score was designed to measure the subject's skill in classifying cues with respect to the problem formulation categories of major importance for the case. In constructing the key for the CUE-PF score, it was necessary to restrict the scope of the key to those problem formulations for which there was sufficient physician data available in order to specify, in a reliable manner, the points to be allotted for the classification of a given

cue under a given problem formulation title. Thus, it was decided to include in the key, as scoring categories, those problem formulations (or groups of similar problem formulations) generated by all six physicians. In some instances, a scoring category consisted of a single problem formulation (e.g., in film 8, diabetes mellitus). In other instances, several formulations generated by the physicians were grouped to form a single category in the key. A category of this latter type was constructed on the basis of three criteria: (1) that the formulations pertain to the same diagnostic "subspace" (e.g., organ system and/or disease mechanism); (2) that the cues listed by the physicians under each formulation be highly similar; (3) that all of the physicians generated at least one formulation belonging to the category. For example, in the film 7 key, several problem formulations (e.g., chronic obstructive lung disease, chronic bronchitis) were grouped to form the scoring category "chronic respiratory problem."

The CUE-PF scoring key consisted of a grid, with cues as one dimension and problem formulation categories as the other dimension. The entry in each cell of the grid was the number of positive or negative points which the subject would obtain for listing the *p*th cue under a problem formulation in the *q*th category. The rationale for the negative points was that if only positive points were awarded a subject could easily attain a perfect score simply

by listing every cue under each of his problem formulation titles. Determination of the sign of the points assigned to each cue X category cell was based on two sources of data: (1) the responses of the physician sample, and (2) independent ratings of the cues (by categories) carried out by the two physician consultants. The latter data were collected in order to reduce the effect of sampling error (in the original physician sample) on the classification of cues as relevant or irrelevant to a problem formulation categories. The primary concern was that negative points be assigned to a cell only if the cue was clearly irrelevant to a problem formulation category, and not simply because the members of the physicians sample had omitted some potentially relevant cue(s) in recording their problem formulations. The following criteria were used to determine the entries in each cue X category cell of the scoring grid:

<u>Cell entry</u> :	<u>Criteria</u> :
+ (CUE points)	cue listed (or rated) as relevant to titles in the category by at least two (sample and/or consulting) physicians
- (CUE points)	cue not listed (or rated) as relevant to any titles in the category by any of the (sample or consulting) physicians
0 points	cues listed (or rated) as relevant to titles in the category by only one (sample or consulting) physician

NOTE: CUE points = the number of points allotted to the cue in the CUE scoring key.

To summarize: the scoring key was designed: (1) to reward the subject for listing relevant cues under a problem formulation title; (2) to penalize him for listing clearly irrelevant cues; and (3) to neither reward nor penalize him for listing cues whose relevance was indeterminant. In addition, rules were incorporated into the key to penalize the subject for various errors (e.g., listing "weight gain" if the patient said he has lost weight; listing a relevant cue but failing to indicate that it was a relevant <u>disconfirmatory</u> cue, as was required by the instructions).

The subject's CUE-PF score consisted of the sum of points he obtained for each cue he listed under any title included in any of the problem formulation scoring categories. The scoring key for film 7 consisted of a 20 cues X 4 categories grid. Summation of points across the grid yielded a score range of -22 to +68. The scoring key for film 8 consisted of a 33 cues X 9 categories grid, and yielded a score range of -122 to +140. Thus, the range of possible scores on the CUE-PF variable (summed over posttest tasks) was -144 to +208. It should be noted, however, that a score in the negative part of the range would be most unlikely to occur.

The CUE-PF score differs from the PF and CUE scores in several ways. First, each of the latter two scores measures a single aspect of the subject's performance: his

problem formulation titles (irrespective of the cues listed under them), or his cue utilization (irrespective of the titles under which they are listed.) The CUE-PF score, on the other hand, measures the way in which the subject classified cues with respect to problem formulation titles. The rationale for this score is the hypothesis that the capacity to categorize cues under multiple problem formulation titles is not a simple linear function of the skills measured by the PF and CUE scores. Second, while the PF and CUE scores are designed to measure both the thoroughness and appropriateness of the subject's performance in each area, the CUE-PF score focuses primarily on the dimension of appropriateness. Unlike the other keys, which each include a fairly exhaustive list of potential responses, the CUE-PF key permits scoring only of those responses that fall within a set of problem formulation categories which were found to characterize the performance of all of the experienced physicians. Thus, if a subject generated a problem formulation that was not included in the scoring categories, the cues associated with this title were not scored.

<u>R-PF score</u>.--This score was designed to measure the degree to which the subject had included, in his tentative assessment, appropriate statements of probable functional relationships between problem formulations. The key for this score included a list of the statements made by the

physician sample, as well as a list of "Other Acceptable Responses" devised in collaboration with the two physician consultants. Points were allotted to statements as follows:

No. of physicians who made the statement	Points
n = 5-6	6
n = 2-4	3
<pre>n = 1 (or judged acceptable     by the two physician     consultants)</pre>	1

The subject's R-PF score consisted of the sum of points obtained for each statement of functional relationships he made, with the number of points obtainable for "Other Acceptable Responses" restricted to a maximum of three (i.e., three one-point statements). The key for film 7 included two statements (plus a list of 9 "Other Acceptable Responses"); the key for film 8 included three statements (plus a list of 4 "Other Acceptable Responses"). Each key yielded a maximum possible score of 12 points; thus, the range of the R-PF variable was 0-24.

### Reliability and Validity

Two aspects of reliability were of concern in the present study: (1) the stability of the posttest scores obtained by independent scorings of the subjects' responses (i.e., inter-scorer reliability); and (2) the generalizability of the posttest scores beyond the sample of two

cases included in the posttest to the domain of potential cases represented by office practice in internal medicine.

Estimation of inter-scorer reliability was based on the sets of scores obtained by having the experimenter and a second professional in medical education independently score the responses of a sample of six subjects. Three subjects were randomly selected from each experimental condition. The response booklets of three subjects (one from each condition) were used to train the second scorer in the use of the scoring keys and instructions. The responses of the remaining six subjects were then scored by the experimenter and by the second scorer. The second scorer was not aware of the experimental condition to which each subject belonged. Interscorer reliability on the variables CUE, PF, CUE-PF and R-PF was computed by means of the Ebel (1951) intraclass correlation formula. A high degree of restriction in range on the R-PF variable resulted in what was judged to be a spuriously low estimate of inter-scorer agreement by means of the intraclass correlation formula. Thus, for this variable an index proposed by Holsti (1969) for use in content analysis was also calculated. The reliability estimates obtained on each variable are presented in Chapter V.

Let us now consider the question of generalizability. On the whole, the issues of reliability and validity have received insufficient attention from researchers in the

field of problem solving. In most studies some consideration is given to devising a problem-solving task which meets certain criteria of "face validity," i.e., a task which presents the subject with a nontrivial problematic situation, and which requires some degree of productive thinking in order to arrive at a solution. But to a large extent the more substantive psychometric issues pertaining to reliability and validity have been ignored: e.g., how reliable are problem-solving scores, given the very small number of tasks typically employed; how valid are inferences about general parameters of human problem solving, given the restricted range of tasks typically employed? An answer to these questions is not easily forthcoming. Problemsolving research poses some particularly difficult psychometric dilemmas. To begin with, it is difficult to define the unit of behavior that is equivalent to the classical test item. It is generally impossible to decompose a problem solution into a set of independently scorable events that would be comparable across subjects. The entire problem may be considered as one "item," but this poses other difficulties. In order to meet even the most informal criteria of face validity the investigator must devise tasks of considerable complexity. This, in turn, severely limits the number of tasks that can be included in any single evaluation instrument.

At the theoretical level, probably the best approach to determining the psychometric properties of a problemsolving test is offered by Cronbach's theory of generalizability (Cronbach, <u>et al.</u>, 1963). Under Cronbach's theory, the classical considerations of reliability and validity coalesce into the consideration of generalizability. The question of central interest in generalizability analysis is to determine the dimensions of the domain to which one can generalize based on a set of empirical observations.

The posttest devised for this experiment was similar in a number of respects to the instruments that have been used in most studies of problem solving, i.e., the subject was given a reasonably complex, problematic task to carry out, but the number of "items" in the posttest was very small. Nevertheless, an attempt was made to give all consideration possible, within the limits of the design of the study, to the issue of generalizability. First of all, in selecting the posttest films consideration was given to choosing two cases that would constitute a reasonably representative sample of the types of problems encountered in the domain of internal medicine. Second, in order to estimate the degree to which it would be possible to generalize from the subjects' scores on the two tasks to the domain of interest, the Cronbach (1951) coefficient alpha formula was used, but modified so as to eliminate between-group differences induced by the experimental

manipulation. The formula used, as well as the resulting estimates for each dependent variable, are presented in Chapter V.

#### Additional Dependent Measures

In addition to the four major dependent variables defined above, a variety of other measures were calculated. These measures included:

 The PF scores of the experimental subjects on the six training tasks;

 The number of items of each type checked on the "Recognition of Cues" task, for subjects under all three conditions;

3. PF, CUE and CUE-PF scores based on the subject's total responses after carrying out the "Additions to Response Sheets" task, for subjects under all three conditions;

4. The experimental subjects' responses to the questionnaire, by item, and summarized in terms of three scores: evaluation of the films (EV FILM), evaluation of the feedback materials (EV FB), and evaluation of the general effectiveness of the training program (EV GEN).

5. Several measures (to be defined) pertaining to the structure of the subject's set of problem formulations on each posttest task.

The purpose of these measures was to aid interpretation of the experimental outcomes of primary interest:

namely, the hypotheses regarding between-group differences as assessed by the PF, CUE, CUE-PF, and R-PF scores on the basic posttest tasks. Thus, the above measures should be regarded merely as supplementary sources of data, of interest primarily as they contribute to an understanding of the experimental outcomes on the four major dependent variables.

#### Hypotheses

The two experimental hypotheses, presented in general terms in Chapter I, will now be operationally defined as follows:

Hypothesis 1:

The average performance of second-year medical students who have received problem formulation training (Treatment I and Treatment II) will be superior to that of students who have not received training (control group), as measured by four dependent variables: (1) CUE score, (2) PF score, (3) CUE-PF score, and (4) R-PF score.

Hypothesis 2:

The average performance of second-year medical students who have received problem formulation training involving outcome and process feedback (Treatment II) will be superior to that of students who have received problem formulation training involving only outcome feedback (Treatment I), as measured by the four dependent variables: (1) CUE score, (2) PF score, (3) CUE-PF score, and (4) R-PF score.

# Analysis

The two experimental hypothesis were tested by means of multivariate and univariate analyses of covariance. In addition, a number of supplemental analyses were conducted.

in order to address questions that have been raised at various points in these chapters, or that were suggested as a result of the outcomes of the hypotheses tests. Chapter V reports the method and results of each analysis undertaken.

#### CHAPTER IV

### ANALYSIS OF THE PHYSICIAN DATA

This chapter presents an analysis and discussion of the data that were collected from the sample of eight experienced physicians. The chapter consists of three sections, each dealing with one of the research questions listed in Chapter I:

- 1. How early in the clinical workup does the physician begin to generate problem formulations?
- 2. What is the structure of a set of initial problem formulations?
- 3. What cognitive processes are involved in the generation of initial problem formulations?

The primary reason for collecting the physician data was to obtain a basis for the development of the materials to be used in the training experiment. However, an analysis of these data is of interest in itself in so far as it may contribute to the elaboration of a theory of medical problem solving. Given the small size of the physician sample, the findings reported in this chapter - must be regarded as highly tentative. But, because the procedure used in this study permits a more in-depth appraisal of initial problem formulation outcomes and

processes than has been forthcoming from previous investigations, the findings may be of value in suggesting hypotheses and questions to be explored in future research.

For each of the eight filmed cases, data were obtained from six (or in one case seven) of the eight physicians. Thus, each analysis reported in this chapter is based on a total of 49 responses. Because of the limited size of the sample, only descriptive statistical analyses were conducted.

## Generation of the First Problem Formulation

It will be recalled from Chapter III (p. 55) that the physician was asked to report what thoughts had come to mind: (a) after the initial 30-second view of the patient, and (b) after having read the nurse's sheet. When problem formulations were not reported at either of these points, the physician was asked, at the end of the film, to attempt to recall the point in the interview at which he first generated a problem formulation.

On the basis of a frequency distribution of these data, it was ascertained that:

 In 17 instances out of 49 (34.7%) the physician generated his first problem formulation after the initial 30-second view of the patient. The formulations generated at this point were of two types: (1) a formulation of "psychological problem" based on the patient's general

appearance and manner, and (2) formulations pertaining to organic disorders, e.g., "respiratory problem," based on specific nonverbal cues, e.g., "the patient's cough."

2. In 4 instances out of 49 (8.2%) the physician generated his first problem formulation after reading the nurse's sheet. In all four instances a formulation of "infection" was generated on the basis of the cue of elevated temperature.

3. In 26 instances out of 49 (53.1%) the physician generated his first problem formulation on the basis of the patient's presenting complaint. These formulations varied across cases depending on the nature of the complaint, but there was a high degree of consistency within each case with respect to the type of formulation generated.

4. In 2 instances out of 49 (4.1%) the physician generated his first problem formulation after the patient's had presented several complaints. Both of these instances occurred on film 1 in which the patient's presenting complaint (fatigue and weakness) was highly general. It was not until the patient mentioned that he also had abdominal pain that, in these two instances, a problem formulation was generated.

Two factors must be borne in mind in attempting to generalize on the basis of these data regarding the earliness with which problem formulations are generated in actual clinical practice. First, the demand characteristics of

the experimental task, as well as the fact that the subject did not have to devote part of his attention to the task of data elicitation (i.e., he obtained data by viewing a film rather than by actively engaging in an interview), may have led the physicians to generate problem formulations somewhat earlier than they would do in actual practice. Second, except in the instances where problem formulations were reported after the initial view of the patient or after the presentation of the nurse's sheet, some degree of retrospective distortion may have affected the physician's report of the point at which he first generated a problem formulation. The findings of this study probably overestimate the earliness with which problem formulation typically occurs in clinical practice. Nevertheless, the fact that in 47 instances out of 49 (95.9%) a first problem formulation was generated no later than one minute into the interview provides evidence that physicians are able to generate problem formulations very early, on the basis of very minimal data, and in actual practice most probably do generate problem formulations relatively early, at least within the first five minutes of the workup.

# The Structure of a Set of Initial Problem Formulations

In Chapter II (p. 32) it was proposed that a set of initial problem formulations defines the dimensions of the functional problem space within which the physician's

search for a diagnosis is conducted. The purpose of this section is to describe the way in which a set of initial problem formulations is structured. Two topics will be dealt with: (1) the features characteristic of a set of problem formulations; (2) the size and organization of a set of problem formulations.

## Structural Features

An examination of the physician data indicated that what results from the physician's information-processing activity during the early part of the workup is not a unidimensional list of problem formulations. Rather it is a <u>structured set</u> of formulations which may be described in terms of four features: (1) hierarchical organization; (2) competing formulations; (3) multiple subspaces, and (4) functional relationships.

1. Hierarchical organization. A set of problem formulations may include formulations that are organized into a general-to-specific hierarchy that pertains to a single diagnostic category (e.g., an organ system, a disease mechanism). For example, a physician may generate a problem formulation, such as "GI disorder" and, as subcategories under this formulation, one or several more specific formulations, such as "inflammatory bowel disease" and/or "intestinal malignancy."

2. Competing formulations. A set of problem formulations may include formulations that provide
alternative explanations for some group of symptoms. For example, a physician may generate "inflammatory bowel disease" and "intestinal malignancy" as competing problem formulations.

3. Multiple subspaces. A set of problem formulations may include subsets of formulations that pertain to different types of diagnostic categories, e.g., different organ systems and/or different disease mechanisms. Each such category may be considered to designate a "subspace" within the functional problem space in which the physician is operating. For example, a physician may generate a set of formulations that consists of four subspaces: (1) "GI disorder," (2) "diabetes mellitus," (3) "anemia," and (4) "cardiovascular problem."

4. Functional relationships. A set of initial problem formulations may include functional relationships which the physician hypothesizes to exist between certain problem formulations. For example, a physician may consider "anemia" to be secondary to "GI disorder."

In order to illustrate the way in which the four features characterize the structure of a set of initial problem formulations, a structural diagram of the composite set of problem formulations generated by the physician sample for one film has been prepared (Figure 3). Although the diagram pictures a more extensive set of problem formulations than would normally be generated by any single individual, it may serve as a useful illustration for the



Structural Diagram of the Composite Set of Problem Formulations Generated by the Physician Sample for Film 1. . m Figure

following commentary on the four features. The number of subspaces indicates the scope (or range) of diagnostic categories included in the problem space. It is the superordinate horizontal dimension of a set of formulations. Subspaces may be competitors (e.g., "GI disorder" versus "diabetes mellitus"); they may be compatible but unrelated (e.g., "diabetes" and "anemia"); they may be functionally related (e.g., "anemia" secondary to "GI disorder"). Some subspaces may consist of a hierarchy of formulations (e.g., "GI disorder" hierarchy) while others may consist of a single formulation that is highly general (e.g., "cardiovascular problem") or very specific (e.g., "diabetes mellitus"). Hierarchical organization of formulations indicates the degree to which the problem space is elaborated on a vertical dimension. Competing formulations may exist within subspaces (e.g., "inflammatory bowel disease" versus "intestinal malignancy"), or between subspaces (e.g., "anemia" versus "cardiovascular problem"). Functional relationships may be hypothesized at the level of subspaces (e.g., "anemia" secondary to "GI disorder"), but can also be hypothesized at the lower levels of subspace hierarchies (e.q., a physician could hypothesize "blood loss anemia" secondary to "ulcerative colitis").

The set of problem formulations generated by each physician for each film was coded in terms of the features it exhibited (see Appendix G). The results of this analysis

are summarized in Table 7, by film (percent of subjects' whose sets of formulations exhibited each feature) and by subject (percent of films for which a subject's set of formulations exhibited each feature).

	Feature						
	Hierarchical Organization	Competing Formulations	Multiple Subspaces	Functional Relationships			
Film	% of s	subjects whose exhibited ea	sets of for ch feature	mulations			
1 2 3 4 5 6 7 8	67 100 100 67 83 57 33 83	100 100 100 100 100 100 100	83 100 33 100 67 100 100 100	83       67         .00       100         33       0         .00       100         67       33         .00       0         .00       100         .00       50			
Subjec	ubject % of films for which a subject's set of formulations exhibited each feature						
A B C D E F G H	38 57 75 75 86 100 75 100	100 100 100 100 100 100 100	75 100 100 75 71 71 100 100	38 71 100 50 43 43 63 50			

TABLE 7.--Features Characteristic of Individual Sets of Problem Formulations, by Film and by Subject.

The following discussion of each feature will consider: (1) the consistency of its occurrence across films and across subjects, and (2) the types of factors which may influence its occurrence.

Of the four features, <u>competing formulations</u> is the only one which is consistently characteristic of all physicians' sets of formulations for all eight films. Thus, the present data suggest that competing formulations is an essential feature of any experienced physician's set of initial problem formulations. It is not surprising that this feature stands out as the most salient characteristic of a set of initial problem formulations. As noted in Chapter II, the entertaining of multiple competing hypotheses is a primary means by which the scientific thinker seeks to avoid the pitfall of becoming prematurely wedded to a favored, but possibly incorrect, hypothesis.

The feature <u>hierarchical organization</u> is present a high proportion of the time for most films and for most subjects. But, it also shows a good deal of variability across subjects and across films. Thus, in contrast to competing formulations, the occurrence of this feature appears to be influenced by both task environment and individual difference variables.

Comments in several of the physicians' recall protocols indicated that, having generated a hierarchy of problem formulations, they would evaluate data subsequently collected with respect to formulations at each level in the hierarchy. It may be hypothesized that a hierarchy of

formulations serves a dual purpose. On the one hand, the early generation of specific formulations would help to guarantee that those cues of particular relevance to the establishment of a differential diagnosis are elicited and interpreted. On the other hand, by continuing to entertain a more general problem formulation category (that subsumes the specific formulations), the physician is more likely to avoid the "blind alley" pitfall that could result if data collection and interpretation were narrowly focused on specific diagnostic hypotheses and these hypotheses were disconfirmed.

A further rationale for the feature of hierarchical organization is suggested by the research literature on the role of organization in memory (e.g., Mandler, 1967; Collins and Quillian, 1969). A hierarchy of problem formulations permits more parsimonious storage of cues. Consider this example.

#### Example:

A physician obtains 8 cues of relevance to X<sub>0</sub> (where X<sub>0</sub> is a relatively general problem formulation, such as "GI disorder"). Of the 8 cues, three cues (#2, #3, #7) are of particular relevance to X<sub>1</sub>, two cues (#2, #6) are of particular relevance to X<sub>2</sub>, and two cues (#2, #8) are of particular relevance to X<sub>3</sub> (where X<sub>1</sub>, X<sub>2</sub>, and X<sub>3</sub> are more specific problem formulations, such as "ulcerative colitis," "intestinal malignancy," and "psychogenic diarrhea.")
If the physician generates a two-level hierarchy of problem formulations, the cues may be stored as follows:



-If on the other hand, if the physician generates a single-level list of specific formulations, the cues must be stored as follows:

х	x	х
1,2,3,4,5,	 1,2,3,4,5,	 1,2,3,4,5,
6,/,8	6,/,8	6,7,8

As the above example illustrates, hierarchical organization increases the number of categories to be stored (from 3 to 4), but greatly reduces the amount of information to be stored within categories (from a total of 24 units to a total of 15 units).

The feature <u>multiple subspaces</u> is present a high proportion of the time. With respect to subjects, this feature shows a relatively restricted range of variability (i.e., it is present for at least 70% of the films viewed by each subject). Thus, individual difference variables appear to have less of an effect on the occurrence of this feature than on the occurrence of hierarchical organization. With respect to films, the range of variability is fairly broad, but the distribution is highly skewed (i.e., for five films all subjects generated formulations in more than one subspace). This would appear to suggest that

for many medical cases (e.g., films 2,4,6,7,8) task environment variables are of primary importance (and thus elicit consistent occurrence of this feature across all subjects) while for other medical cases (e.g., films 1,3,5) task environment variables are less powerful and individual difference variables may play more of a role. There are two task-related factors which probably contribute to the high frequency with which this feature typically occurs. One is the fact that a patient may often have more than one medical disorder. Thus, resolution of such cases would require that multiple disease mechanisms and/or disorders in multiple organ systems be considered. A second factor is the ambiguity of the cues obtained during the early part of the workup: many cues are inherently nonspecific (e.g., weakness and fatigue), and even relatively specific cues (e.g., substernal pain) may be compatible with multiple disease mechanisms and/or multiple organ system involvements.

The feature <u>functional relationships</u> is more likely to be absent from a set of initial problem formulations than any of the other three features. There is a considerable degree of variability across subjects with respect to this feature, with some subjects showing a much greater tendency to hypothesize functional relationships than others. With respect to films, the data appear to suggest that for many cases (e.g., films 2,3,4,6,7) task environment variables are powerful enough to elicit consistent outcomes across all

physicians (i.e., either all subjects hypothesized functional relationships or none did), while for other cases (e.g., films 1,5,8) the occurrence of this feature is likely to vary and may be largely a function of individual difference factors. One task environment variable which appears to influence the occurrence of this feature is the age of the patient: several physicians noted when the patient is older than forty (e.g., films 2,4,7) he is more likely to have multiple disorders that are functionally related.

## Size and Organization

The size of a set of initial problem formulations may be measured in terms of two variables: (1) the number of problem formulations it contains, and (2) the number of subspaces it contains (see Appendix G). Table 8 presents the mean and range on these variables, by film (across subjects) and by subject (across films). For films, the average number of problem formulations ranged from 3.5 to 8.8, and the average number of subspaces from 1.3 to 5.0. For subjects, the average number of problem formulations ranged from 3.6 to 7.8, and the average number of subspaces from 2.6 to 4.3.

The two measures of the size of a set of problem formulations are highly correlated: a product-moment coefficient of .70 with film as the unit, and a coefficient

-	Number	of	Number	of
	Problem For	mulations	Subspa	ces
	Average	Range	Average	Range
Film				
1	4.8	3-7	3.2	1-5
2	8.2	5-14	5.0	3-6
3	3.5	3-6	1.3	1-3
4	5.0	3-7	3.7	3-6
5	4.3	3-7	2.0	1-3
6	6.9	4-12	3.4	2-3
7	4.7	3-7	3.8	3-5
8	8.8	4-11	4.3	3-5
Subject				
A	3.6	3-5	2.6	1-4
B	6.0	4-11	4.3	2-6
C	6.5	3-9	3.8	3-5
D	5.3	3-7	2.8	1-4
E	5.3	3-7	2.6	1-5
F	5.7	3-11	2.9	1-5
G	7.8	3-14	4.3	2-6
H	7.0	3-10	3.5	1-6

TABLE 8.--Number of Problem Formulations, and Number of Subspaces: Average and Range by Film, and by Subject.

of .78 with subject as the unit. These correlations indicate that the two measures have a very sizable proportion of variance in common. Thus, the question arises as to the rationale for considering that the two measures pertain to distinct psychological entities. The rationale for this distinction derives from an evaluation of the data in Table 8 in terms of the research literature on the role of organization in memory.

Research by Mandler (1967) has indicated that a subject typically organizes and stores items in terms of (5±2)

categories. Moreover, there is some evidence (Wortman, 1972; Wortman and Kleinmuntz, undated) that Mandler's (5±2) parameter applies to the information-processing behavior of physicians. An examination of Table 8 reveals that the number of initial problem formulations a physician generates may in some instances considerably exceed the storage capacity of working memory (e.g., films 2,6,8; subjects B,F,G,H). However, the number of subspaces generated for a given case, or by a given subject, never exceeds six. Thus, it would appear that however many problem formulations a physician generates, the maximum number of subspaces into which these formulations are grouped is consistent with the upper bound of the parameter that has been found to govern the storage of information in working memory. This finding would seem to attest to the psychological reality of the subspace as the superordinate unit in a set of problem formulations.

How many subspaces a physician generates (within the limit imposed by memory capacity) is probably a function of both individual difference variables (e.g., his knowledge of relevant medical content) and task environment variables (e.g., the complexity of the case). It was possible to identify one task environment variable which appears to have influenced the performance of this physician sample: namely, the number of different organ systems to which the patient's complaints pertained. When

this variable is correlated with the minimum number of subspaces generated for each film, a product-moment coefficient of .72 is obtained. Thus, it would appear that while memory capacity imposes a limit on the maximum number of subspaces a physician generates, the number of organ systems to which the patient's complaints pertain is one task environment variable that governs the minimum number of subspaces generated.

Let us now consider the way in which problem formulations are organized into subspaces. Examination of the physician data reveals that problem formulations are never evenly distributed across subspaces. In the typical case, e.g., a problem space containing two to five subspaces, there are usually several subspaces containing only one problem formulation, and several other subspaces containing from two to at most four problem formulations at the same level of specificity. The subspaces that are hierarchically elaborated typically include only two (or at most three) levels of specificity. Thus, the number of units included in a subspace never exceeds, but in many instances does fall considerably below, the (5±2) parameter proposed by Mandler. There are several factors which may account for this finding. (1) In some instances, the subspace category may be at a level of specificity which does not admit further hierarchical elaboration of diagnostic relevance (e.g., the subspace "diabetes mellitus" in Figure 3).

(2) In other instances, it would be possible to generate a hierarchy of subordinate formulations, but the current data base is so limited with respect to that subspace that further hierarchical elaboration would be fruitless (e.g., the subspace "cardiovascular problem" in Figure 3).

## Conclusions

On the basis of the preceding analyses, several tentative conclusions may be proposed regarding the structure of a set of initial problem formulations:

1. The subspace is the superordinate unit in a set of problem formulations. Typically, there are two to five such units.

In the typical case, some subspaces contain
 2-4 hierarchically organized formulations, while other subspaces, contain only a single formulation.

3. In virtually every case, there are competing formulations at the level of subspace categories and/or at the level of specific formulations within subspaces.

4. In some cases, there may also be functional relationships linking subspaces and/or specific problem formulations.

# Processes Involved in Generating Initial Problem Formulations

As described in Chapter III (p. 56), two types of data relevant to problem formulation processes were collected: (1) retrospective recall data, (2) process checklist data. This section will present findings that were derived from an analysis of each type of data.

For each film the subjects' recall protocols were collated by means of a process summary sheet. This sheet summarized, in outline form, the time sequence of mental events that were reported by the physicians, including a notation of the number of subjects reporting each event.

A review of the process summary sheets for all eight films yielded several observations regarding the processes underlying the generation of initial problem formulations. The discussion of these observations will be organized so as to indicate how problem formulation processes are related to each of the structural features described in the first section of this chapter.

Generation of a hierarchy of problem formulations.---Kleinmuntz (1968; Wortman and Kleinmuntz, undated) has proposed that the diagnostic process is characterized by hierarchical search which proceeds from general problem formulation categories to increasingly specific diagnostic formulations. The data from the present study indicate that a physician's initial problem formulations cannot be characterized as either highly general or highly specific. In fact, a set of initial problem formulations typically includes hierarchies of formulations at various levels of specificity. Moreover, the data from the physicians' recall protocols indicate that the elaboration of a problem

formulation hierarchy may proceed in three ways: (1) from general to specific; (2) from specific to general; (3) generation of general and specific formulations virtually simultaneously. Each of these processes may be illustrated by examples from the recall protocols.

Example of process 1: general to specific In viewing film 1, nearly all physicians generated the general formulation of 'GI disorder' on the basis of the patient's complaint of abdominal pain. Subsequently, when the patient mentioned having diarrhea (with blood and mucous in his stools), they generated the more specific formulations of 'ulcerative colitis' and/or 'regional enteritis.'

Example of process 2: specific to general In viewing film 6, nearly all physicians generated the specific formulations of 'rheumatoid arthritis' and/or 'ankylosing spondylitis' relatively early on the basis of certain cues (i.e., stiff back in a.m., but loosens up during the day; back pain plus dyspnea). Subsequently, when it was learned that the patient also has knee and ankle inflammation, they generated the more general formulation of 'polyarthritis.'

Example of process 3: general and specific simultaneously In viewing film 5, nearly all physicians generated the general formulation of 'acute infectious illness' very early on the basis of the cues fever, headache and sleepiness of three days duration. They noted that almost simultaneously they thought of several specific types of infectious illnesses (e.g., 'viral flu,' 'infectious mononucleosis,' 'infectious hepatitis') that are highly prevalent in a college student population.

To summarize: it is necessary to distinguish between the processes of generating a problem formulation hierarchy, and the product of these processes. While the product may be represented as a general-to-specific hierarchy of formulations, the process of generating the hierarchy may take one of three forms. For a given case, however, there was

a substantial degree of consistency (across subjects) in the process reported. Thus, it would appear that task variables, rather than individual difference variables, were the major determinants of which process was employed.

Generation of competing formulations.--The recall protocols provided evidence of two types of processes underlying the generation of competing formulations: (1) generation of competitors at a single point in time on the basis of the same set of cues; (2) generation of competitors over several points in time on the basis of different cues. Examples from the recall protocols will serve to illustrate each of these processes.

Example of process 1:

In viewing film 1, nearly all physicians generated a list of competing formulations (e.g., 'ulcerative colitis,' 'regional enteritis,' 'intestinal malignancy') at almost the same point in time, on the basis of the patient's report of diarrhea with blood and mucous.

Example of process 2:

In viewing film 8, all physicians generated the formulation of 'GI disorder' early in the interview on the basis of the cues abdominal pain and vomiting. Much later, they generated the formulation of 'diabetes,' as a competitor to 'GI disorder,' on the basis of the patient's report of increased appetite, thirst and urination.

It is probable that the associative mechanisms underlying the above processes are quite different. In the case of the first process, we may hypothesize two types of underlying associative mechanisms: (1) association from cue(s) to a list of competing formulations; (2) association from cue(s) to one formulation, and from this formulation to another competing formulation, and so on. In the case of the second process, we may hypothesize an associative mechanism of the following sort: association from one set of cue(s) to a formulation; association from another set of cue(s) to another formulation; associative link-up of the two formulations as competitors. As was the case for hierarchical problem formulations, diverse associative processes may result in the same product, i.e., a set of competing formulations to be stored in memory.

Generation of multiple subspaces.--Evidence from the recall protocols indicated that there are two types of processes underlying the generation of multiple subspaces: (1) generation of multiple subspaces at a single point in time on the basis of the same set of cues, and (2) generation of multiple subspaces at several points in time on the basis of different cues. Examination of the recall data revealed that there are several task variables which appear to govern the generation of multiple subspaces. The first process listed above generally occurred under two circumstances: (a) when the patient's complaint was of a general or multisystem nature, and (b) when the location of a specific complaint indicated that several organ systems could be involved. The second process listed above generally occurred when complaints, reported at different points on the interview, pertained to different organ systems or implied different disease mechanisms. These generalizations may be illustrated by the following examples.

Examples of process la:

In film 1 the patient's presenting complaint of fatigue and weakness is highly general and could be compatible with diverse organ system involvements and/ or disease processes. Thus, some physicians generated formulations belonging to two distinct subspaces: (1) 'anemia,' and (2) 'cardiovascular problem.'

Examples of process lb:

In film 2 the patient indicates a substernal location of her chest pain. On the basis of this cue, all physicians generated formulations belonging to two subspaces: (1) 'cardiac problem,' and (2) 'upper GI problem.'

Example of process 2:

In film 7 the patient's presenting complaints of two days duration led all physicians to generate formulations belonging to the subspace 'acute respiratory infection.' Subsequently, when the patient reported symptoms of several years duration, they generated formulations in two additional subspaces: (1) 'chronic respiratory problem,' and (2) 'cancer.'

When multiple subspaces were generated by processes la and lb they were usually competitors, whereas multiple subspaces generated by process 2 most often pertained to disorders which were not mutually exclusive (i.e., some sort of functional relationship between subspaces, was hypothesized, or the subspaces pertained to concommitant but unrelated disorders).

<u>Generation of functional relationships</u>.--The data from the recall protocols indicate that, as each new problem formulation is generated, the physician generally considers how it might be functionally related to the formulations he has previously generated. Thus, functional relationships between problem formulations are usually not hypothesized until the physician has generated at least two noncompeting formulations. However, this is not always the case. In viewing film 2, for example, several physicians noted that because of their initial impression of the patient (i.e., a middle-aged, obese, anxious appearing woman), they expected from the outset that she might have multiple, interrelated problems of both an organic and psychological nature.

Having presented and discussed the findings from the analysis of the recall protocol data, let us now consider the findings from the analysis of the process checklist data. As indicated in Chapter III (p. 54), the items in the process checklist pertained to the following aspects of the act of generating initial problem formulations:

- 1. modes of mental representation;
- 2. strategies of problem formulation, including,
  - a. initial routines,
  - b. general strategies;
- 3. associative processes of problem formulation;
- 4. cue utilization.

The classification of items according to the above topics is presented in Table 9 (p. 135).

The analysis of the checklist data was designed to determine, for each item: (a) its overall importance as a characteristic of the act of generating initial problem formulations, (b) its stability with respect to subjects (across tasks), (c) its stability with respect to tasks (across subjects). The first step in the analysis was to construct a subject X task (i.e., film) data matrix for each item. In the 49 matrix cells for which data were available, a 1 was entered to indicate that the *n*th subject checked the item on the tth task.

The overall importance of an item as a characteristic of the act of generating problem formulations was measured by the relative frequency with which it was checked: i.e., (the number of cells in the item matrix with an entry of 1)/49. The results of these calculations are presented in Table 9.

The subject stability and task stability of each item was measured in terms of the following criteria:

- 1. subject stability: an item was considered to be a stable characteristic of a subject's performance if the item was checked for all but one of the films he viewed;
- 2. task stability: an item was considered to be a stable characteristic of performance on a given task if it was checked by all but one of the subjects who had viewed the film.

A 1 was entered in the margin(s) of the item matrix for each subject, or task, which met the stability criteria defined above.

In order to determine the proportion of cell entries which could be accounted for by using the stability criteria defined above, the following formula was employed:

<del>11</del>	Category	Item No. and Description	Rel. Freq.
I.	Modes of Mental	l. mental list	.84
	Representation	7. mental imagegeneral	.49
		21. mental imageanatomical	.45
		patient	.29
II.	Strategies of Problem Formu- lation		
	A. Initial	16. organic vs. psychogenic	.22
	Routines	10. assume organic	.63
		18. acute vs. chronic	.47
		11. localize organ system	.59
	B. General	12. incidence	.73
	Strategies	19. incidence, plus complaints	.71
		2. seriousness	.45
		4. pathophysiological	• • •
		processes	.35
		17. convergence	.24
		9. divergence(1)	.35
		24. divergence(2)	.69
		3. guick "rule outs"	.43
			• • •
III.	Associative Processes of	8. associationsalient cue	.84
	Problem Formu-	of cues	. 82
	lation		••=
<b>—</b>	-		<b></b>
IV.	Cue	25. focus on verbal cues	.61
	Utilization	20. focus on nonverbal cues	. 39
		6. impression of patient	.71
		13. presenting complaintmore	2
		weight	.43
		14. selective focus on cues	.24
		22. Interretate cues	00
		progressively	.00
		5. STORE CUES, INTERRELATE	65
		Talei	•05

TABLE 9.--The Relative Frequency with which Each Process Checklist Item was Checked.

- where  $N_t =$  the total number of cells in the item matrix, i.e., 49
  - Ne = the number of cells in the item matrix
    whose entries deviated from those that
    would be predicted on the basis of the
    entries in either matrix margin (i.e.,
    a cell entry of 1, but no entry in either
    the subject or task margin; or, conversely,
    no cell entry, but a 1 in either the
    subject or task margin).

The coefficients for each item calculated according to this formula ranged from 65.3% to 89.8%, with an average of 81.8% across all 25 items. Thus, in general, the criteria adopted for measuring subject and task stability accounted for a very large percentage of the observed responses.

In order to summarize the data on item stability with respect to subject and tasks, each item was classified along two crossed dimensions: (1) degree of subject stability: i.e., the number of subjects for whom the item was a stable characteristic of performance across tasks, and (2) degree of task stability: i.e., the number of tasks for which the item was a stable characteristic of performance across subjects. Table 10 presents the results of this classification.

The results of the analysis of the checklist data, presented in Tables 9 and 10, will now be discussed with respect to each of the topics listed on page 133.

	Degree of Task Stability <sup>a</sup>									
		8	7	6	5	4	3	2	1	0
	8		8							
	7		1		15					
Degree of	6		22							
Subject Stability <sup>b</sup>	5						6			
	4				10	12,19 24		5		
	3							11,25		
	2								2,13	3,7 9,18
	1								21	14,20 23
	0									4,16 17

TABLE 10.--Classification of Checklist Items on Two Dimensions:
 (a) Degree of Subject Stability, and (b) Degree of
 Task Stability.

<sup>a</sup>Number of tasks on which the item was a stable characteristic of performance across subjects.

<sup>b</sup>Number of subjects for whom the item was a stable characteristic of performance across tasks.

NOTE: Entries in the cells are the item numbers (see Table 9).

Modes of mental representation.--This topic was concerned with two modes of mental representation: (1) the verbal, and (2) the figural. Four checklist items (numbers 1, 7, 21 and 23) were of relevance to this topic. The data in Tables 8 and 9 suggest the following conclusions regarding the relative importance of verbal versus figural modes of mental representation in generating problem formulations.

The generation of mental lists of problem formulations (item 1) occurred a very high proportion of the time (.84), and was found to be a stable characteristic of nearly every subject's performance on nearly every task. The generation of mental images occurred considerably less frequently (.49, .45 and .29, for items 7, 21 and 23, respectively). Moreover, the occurrence of mental images showed a very low degree of stability with respect to subjects or tasks: there were only two subjects who consistently reported mental images, and only one task for which mental images were consistently reported. Examination of the relative frequency scores for the three mental image items, revealed that when images are generated, they generally pertain to the anatomical location of the patient's problem (item 21), and, less frequently, may consist of an evocation of a previous patient (item 23). In sum, we may conclude that the generation of problem formulations is typically carried out in a verbal mode of mental representation, but that for a few individuals, or occasionally

for all individuals on certain tasks, mental imagery accompanies the predominantly verbal train of thought.

Strategies of problem formulation: initial routines.--This topic was concerned with the occurrence of what may be termed "initial routine strategies" for the generation of problem formulations. The items of relevance to this topic were designed to determine whether one of the physician's first steps in the problem formulation process was: (1) to consider the patient's complaint(s) in terms of an organic versus psychogenic distinction (items 10 and 16); (2) to consider the patient's complaint(s) in terms of an acute versus chronic distinction (item 18); (3) to localize the patient's complaint(s) in terms of an organ system (item 11). All three of these strategies pertain to highly general principles of problem formulations. Thus, this topic is also concerned with whether the physician begins the process of generating problem formulations at a high level of generality. The data in Table 9 and 10 suggest the following conclusions regarding initial routine strategies.

Examination of the relative frequency scores reveals that the physician does not generally begin by trying to make an organic versus psychogenic destination (item 16, .22); most often he tends to assume that the patient's problem is organic (item 10, .63). Item 16 was not a stable characteristic of any subject's performance of performance

on any task. On the other hand, item 10 was consistently checked by four subjects (across tasks), and on five tasks (across subjects). The relative frequency scores of item 18 (.47) and 11 (.59) indicate that these two routines occurred with more frequency than the first routine (item 16), but, like the first routine, showed little or no subject or task stability. In sum, we may conclude that initial routines involving highly general distinctions are not typically the first step(s) in the process of generating problem formulations. Only a few individuals consistently follow such routines, and there are few tasks which consistently prompt their use.

General strategies of problem formulation.--In contrast to the previous topic which was concerned with the physician's initial strategies, this topic is concerned with the strategies that the physician may use throughout the entire 4-6 minute encounter. Four items included under this topic sought to determine whether the physician follows a strategy based upon: (1) consideration of disease incidence (items 12 and 19); (2) consideration of disease seriousness, i.e., its life-threatening implications (item 2); (3) consideration of pathophysiological mechanisms (item 4). Three of the items sought to determine whether the physician follows a convergent strategy, i.e., attempts to come up with one problem formulation that will account for all the data (item 17), and/or either of two

divergent strategies: (1) a "brainstorming" strategy of attempting to think of as many formulations as possible that fit the cues (item 9), or (2) a more modest divergent strategy of attempting as each formulation is generated to think of other possible formulations (item 24). There was also one item (number 3) designed to determine whether the physician had already, within the 4-6 minute interview, performed some quick "rule outs" of certain problem formulations. The data relevant to these items suggest the following conclusions.

Consideration of disease incidence is relatively more important than consideration of disease seriousness in determining the type of problem formulations a physician generates (relative frequency scores of .71 and .73 for the incidence items versus .45 for the seriousness item). Consideration of incidence was consistently checked by four physicians, and on four tasks, while consideration of seriousness was consistently checked by two physicians, and on only one task. Thus, we may conclude that although the physician may not give consideration to either of these factors, he is relatively more likely to direct his search toward diseases of high incidence (which are usually not very serious) than toward diseases of great seriousness (which usually have low incidence).

The item pertaining to consideration of pathophysiological processes (17) was checked a relatively small

proportion of the time (.35). Since knowledge of pathophysiological processes is considered to be one of the foundations of clinical medicine, this result is somewhat surprising. It may be that in the experienced physician the utilization of such knowledge is so well established (routinized) that he is no longer consciously aware of its use in generating problem formulations. On the other hand, it is also possible that the generation of problem formulations is essentially a cue-to-disease associative mechanism which does not require consideration of the pathophysiology underlying diseases processes. This latter hypothesis receives some support from data to be discussed under the topic "associative processes of problem formulation."

Examination of the data for the items on convergent versus divergent strategies of problem formulation reveals that the convergent item was checked quite infrequently (item 17, .24), while one divergent item was checked relatively frequently (item 24, .69) and the other was not (item 9, .35). The convergent item was not consistently checked by any subjects, or on any task. The data for the divergent items indicate that item 24 was consistently checked by four subjects, and on four tasks, while item 9 was checked consistently by only two subjects, and on none of the tasks. It is not surprising that experienced physicians rarely follow a convergent strategy of problem formulation during the initial 4-6 minutes of the workup.

To do so would entail the risk of premature closure: i.e., acceptance of a formulation which may be intellectually appealing (because it can account so parsimoniously for the available data), but possible incorrect. Divergent strategies of problem formulation would of course help to counteract any tendency toward premature closure. Of the two divergent strategies that were considered, one (the strategy of attempting each time a formulation is generated to think of other formulations) was consistently employed by a sizable number of physicians, and on a sizable number of tasks, while the other (a "brainstorming" strategy of attempting to think of as many causes of the patient's symptoms as possible) was not. The less frequent use of the second divergent strategy may be due to the risks which brainstorming could entail: e.g., information overload taxing the capacity of working memory; inefficient data collection to test numerous potential but impulsible hypotheses.

The item on quick "rule outs" was checked with a relative frequency of .43. However, it was checked consistently by only two subjects, and was not consistently checked for any tasks. Thus, we may conclude that while a few physicians consistently rule out some problem formulations generated during the first 4-6 minutes of a workup, most physicians retain all formulations they generate as components of their initial problem space.

Associative processes of problem formulation, -- The items under the two previous topics were designed to determine whether the physician attempts to follow various strategies of problem formulation. The items under this topic were designed to determine whether the act of generating problem formulations entails associative processes, i.e., rapid cue-to-problem formulation retrieval, essentially outside the realm of conscious search. There were two items of relevance to this topic. They sought to determine whether problem formulations were immediately brought to mind: (1) by some "particularly salient cue" (item 8), and/or (2) by a combination of cues (item 15). Both of these items were checked a very high proportion of the time (.84, and .82, respectively). One of the items (8) was checked consistently by all eight subjects, and on seven of the tasks, thus showing a higher degree of subject and task stability than any other item on the checklist. The other item (15) was checked consistently by seven subjects, and on five tasks; thus, it was among the top four items with respect to degree of subject and task stability.

When the data for these items are compared with the data on problem formulation strategies, we are led to conclude that the generation of problem formulations is largely an associative process. Search strategies are employed, more or less frequently depending on the individual

and the case, but appear to be adjuncts to the primary process of associative retrieval. Moreover, the finding that generation of problem formulations is more consistently based on single salient cues, than on combinations of cues, would appear to indicate that the physician's long-term storage of potential problem formulations categories (i.e., disease processes) may be indexed in terms of a very small number of pathognomonic cues for each category, rather than in terms of a complex system of multiple-entry, crossreferenced cues. In sum, the checklist data tend to support the notion that the generation of diagnostic problem formulations consists primarily of rapid cue-to-problem category associative retrieval. As Barrows and Bennett (1972) noted in their study of neurologists, hypotheses seem to literally "pop" into the head of the clinician.

<u>Cue utilization</u>.--The items under this topic were designed to measure several aspects of the physician's behavior with respect to detecting, interpreting, and utilizing cues: (1) whether he focuses on verbal cues (item 25) or nonverbal cues (item 20); (2) whether he gives more weight to the patient's presenting complaint then to subsequent cues (item 13); (3) what role his initial impression of the patient plays in interpretation of the reliability of cues (item 6); (4) whether he focuses his attention on certain cues and pays less attention to others (item 14); (5) whether he attempts to look for relationships among

cues progressively as each cue is presented (item 20), or stores cues and attempts to interrelate them after obtaining data on each major complaint (item 5). The data for these items suggest the following conclusions regarding cue utilization.

First, although in many instances the physician does not selectively focus on either verbal or nonverbal cues, there is a greater tendency to focus on verbal than on nonverbal cues (relatively frequency scores of .61 and .39, respectively). Focusing on nonverbal cues (item 20) was a stable characteristic of only one subject (and no tasks), while focusing on verbal cues (item 25) was a stable characteristic of three subjects (and two tasks). However, the data for item 6 indicated that physicians generally do make use of nonverbal data to form an early impression of the patient (his personality, intelligence, background, etc.), which is used as basis for judging the accuracy and objectivity of the symptoms the patient reports. This item had a relative frequency score of .71, and was consistently checked by five subjects, and on three tasks. Thus, it appears that some physicians make greater use of their initial impression of the patient than others; and that some patients provide more of a basis for doing so (e.g., have more salient nonverbal characteristics) than others.

The items dealing with giving more weight to the patient's presenting complaint than to other cues (item 13), and with giving more attention to certain cues than to others (item 14) had moderate to low relative frequency scores (.43 and .24, respectively) and were consistently checked by only one or two subjects, and on only one (or none) of the tasks. Several physicians' comments in their recall protocols suggest a reason why the physician does not typically give more weight to the patient's presenting complaint: they noted that with some patients there may be a "hidden agenda" of medical problems which must be uncovered by careful questioning, and which may prove to be more important than the presenting complaint.

The two items dealing with the manner in which the physician attempts to interrelate cues were both checked relatively often (item 22, .88; item 5, .65). In constructing the checklist, these two items were considered to describe two contrasting strategies for dealing with cues. However, the data reveal that four subjects consistently checked both of them. Several subjects' comments in discussing the checklist items with the experimenter provided an indication as to why this occurred. They noted that although they do not consciously adhere to either of the strategies described by the two items, both items do refer to aspects of their processing of cues and were therefore checked. They noted that it is not a matter, as the items imply, of

"trying" (or "not trying") to relate each new cue to previous cues; rather, they suggested, relationships among cues simply "come to mind," usually in a progressive manner as each new cue is obtained, but sometimes at a later point in time.

### Conclusions

Several tentative conclusions may be drawn from the analysis of the recall protocol and checklist data regarding the processes involved in the act of generating initial problem formulations.

1. The mode of mental representation involved in the generation of problem formulations is, for all physicians, predominantly verbal. For most individuals mental imagery occasionally occurs, and for a few individuals such imagery consistently occurs. But, for all individuals, mental imagery appears to be an adjunct to the primary verbal mode of representation.

2. Although physicians show some tendency to focus on verbal cues, they do make use of nonverbal cues. One major use of such cues, consistently reported by over half of the physicians, is to form a general impression of the patient that will aid him in judging the accuracy and objectivity of the cues reported verbally by the patient.

3. The generation of problem formulations appears to be primarily a process of direct associative retrieval, rather than one of strategy-guided search. The checklist

data indicated that only two strategies were consistently employed by at least half of the physicians: (1) focusing on diseases of high incidence for the patient's demographic group, and (2) attempting to think of alternatives (or competitors) to each formulation generated. On the basis of the recall protocol data, it was possible to identify various processes underlying the generation of each of the four structural features of a set of initial formulations. In general these processes appeared to be governed by the effect of various task variables on associative retrieval, rather than by consistent use of strategies on the part of the subject. In particular, it may be noted, the data failed to support Kleinmuntz's (1968) notion that the physician follows a "general-to-specific" strategy, at least so far as the generation of initial problem formulations is concerned. To summarize: both the recall protocol and checklist data suggest that the physician's information-processing activity during the early part of the workup consists primarily of associative retrieval of problem formulation labels on the basis of cues, and that this process is mediated to only a limited extent by search strategies.

#### CHAPTER V

## RESULTS OF THE TRAINING EXPERIMENT

This chapter presents the results of the training experiment conducted with second-year medical students. It includes two major sections: (1) results of the analyses conducted to test the experimental hypothesis; and (2) results of several supplemental analyses conducted to aid in interpreting the outcomes of the hypothesis tests.

## Tests of Experimental Hypotheses

### Reliability Estimates

Two types of reliability estimates were calculated on the basis of the posttest data: (1) estimates of interscorer reliability in employing the scoring keys for the CUE, PF, CUE-PF and R-PF variables; and (2) coefficients of generalizability for the CUE, PF, CUE-PF and R-PF variables.

Inter-scorer reliability.--A random sample of six subjects' posttest responses were scored independently by the experimenter and a second person. Since the experimenter would ultimately score all subjects' posttest responses, this procedure was not undertaken in order to estimate the effect of inter-scorer variability (as a source
of error) on the final set of scores. Rather it was carried out as a means of determining the objectivity of the scoring keys that had been devised for each variable. A basic judgmental operation was required in utilizing each key: namely, a judgment as to whether a given component of the subject's response was equivalent to some item in the scoring key. Once such a judgment had been made, the other operations in utilizing the key were largely mechanical. Thus, it was the consistency of these judgments that was of primary concern in estimating inter-scorer reliability.

, The intraclass correlation coefficient (Ebel, 1951) was used to estimate the inter-scorer reliability of six subjects' scores on the variables CUE, PF, CUE-PF and R-PF. These coefficients are presented in Table 11. As shown in the table, the coefficients for the variables PF, CUE and CUE-PF are nearly 1.0. The coefficient for the R-PF

Variable	Intraclass Correlation	Index of Agreement
CUE	.97	
PF	.99	
CUE-PF	.97	
R-PF	.09	.83

TABLE 11.--Inter-Scorer Reliability Coefficients on the Variables CUE, PF, CUE-PF and R-PF.

variable, however, was quite low. Inspection of the data revealed that this was due to a high degree of restriction in range on this variable, rather than a substantial degree of divergence between scorers. It was therefore decided to employ a formula proposed by Holsti (1969, p. 140) for estimating percent of agreement in conducting a content analysis. Since the determination of a subject's R-PF score was based on an analysis of the text of his tentative assessment, a content analysis index was deemed to be appropriate for estimating inter-scorer agreement on this variable. As indicated in Table 11, the index of interscorer agreement on the R-PF variable was very high. Thus, it may be concluded that on all four of the dependent variables scorer would not constitute a source of unreliability in the data.

Generalizability coefficients.--In order to determine the degree to which it would be possible to generalize from subjects' performance on the two posttest tasks to a hypothetical population of randomly equivalent tasks in the domain of internal medicine, generalizability coefficients were calculated for each dependent variable on the basis of all 48 subjects' posttest scores. These coefficients were calculated by means of Hoyt's (1941) analysis of variance technique (which is equivalent to Cronbach's (1951) coefficient alpha). However, because the coefficients were being calculated on experimental posttest data, the

formula was modified so as to exclude between-group variation. This was necessary in order that estimation of the generalizability of the instrument would not be contaminated by treatment effect. The formula used was as follows:

$$r = \frac{MS_{S:G} - MS_{TS:G}}{MS_{S:G}}$$

where MS<sub>S:G</sub> = mean square for subjects within groups MS<sub>--</sub> = mean square for subjects X task inter-

MS<sub>TS:G</sub> = mean square for subjects X task interaction within groups

The obtained coefficients are presented in Table 12. This table also presents the within-group correlation coefficients between scores on the two posttest tasks. As shown in the table, the generalizability coefficients ranged from .48 to .73. Given the fact that the posttest included only two tasks, and that the tasks were selected so as to represent very different cases (with respect to type of medical complaints and patient demographic characteristics), the magnitude of the generalizability coefficients is quite substantial. In fact, as compared to coefficients obtained by other investigators in the field, the coefficients are very high. Lewy and McGuire (1966), for example, obtained coefficient alphas of .35, .27, .21 and .10 for proficiency scores on tests including two tasks (i.e., Patient Management Problems) in each of four domains.

Variable	Generalizability Coefficient	Correlation Between Tasks
CUE	.73	.67
PF	.66	.50
CUE-PF	.55	.41
R-PF	.48	.31

TABLE 12.--Generalizability Coefficients, and Within-Group Correlation Coefficients (Between Tasks) on the Dependent Variables CUE, PF, CUE-PF and R-PF.

In sum, given the very high degree of inter-scorer reliability, and the substantial magnitude of the generalizability coefficients, it was concluded that the posttest instrument had sufficiently strong psychometric properties to permit detection of treatment effects by inferential hypothesis tests.

# Results of Hypothesis Tests

The experimental hypotheses were tested in the following manner. First, a multivariate analysis of covariance (MANCOVA) was conducted in order to determine whether there was a significant main effect for treatment as measured by the set of four dependent variables. Second, stepdown F ratios obtained from the MANCOVA, as well as univariate F ratios obtained from an analysis of covariance (ANCOVA) on each dependent variable, were tested in order to identify the dependent variable(s) on which a significant treatment effect occurred. Third, for each variable having a significant univariate F ratio, the Scheffé <u>post hoc</u> confidence interval procedure was used in order to test for significant differences between each pair of experimental conditions. Although the experimental hypotheses were stated in directional form, an omnibus nondirectional test of the null hypothesis for treatment was conducted in order that significant differences in either the predicted direction or the opposite direction would be detected through <u>post hoc</u> comparisons. Both the MANCOVA and the ANCOVAS were conducted using the Finn (1970) program on the Michigan State CDC 6500 Computer. The Scheffé <u>post hoc</u> comparisons were computed by the experimenter using the formulas presented in Glendening (1973).

For one subject a score on the covariate (Focal Problems Exam) was not available. Thus, prior to the analysis, this subject's score was estimated by means of the following regression equation:

> $\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4$ Where  $\hat{Y}$  = estimated Focal Problems Exam score  $X_1$  = PF score  $X_2$  = CUE score  $X_3$  = CUE-PF score  $X_4$  = R-PF score

The estimated exam score for this subject was then used in the analysis.

The observed means and standard deviations on the four dependent variables and on the covariate for each experimental condition are presented in Table 13. The means on the dependent variables adjusted for their relationship to the covariate are presented in Table 14.

Examination of the means in Table 14 reveals that on the variables CUE, PF and CUE-PF the Treatment I means are consistently higher than the Treatment II means, and the latter are consistently higher than the control means. On the variable R-PF, the means for Treatment II and control are very close, while the mean for Treatment I is higher than the other two. Examination of the standard deviations in Table 13 indicates a higher degree of variability under the control condition than under the treatment conditions, especially on the variable CUE-PF.

The analysis of covariance fixed effects model is based on the following assumptions: (1) independence of observations on the dependent variable; (2) normality of the conditional population distributions; (3) equality of the conditional population variances; (4) equality of the population regression slopes. Empirical studies (reviewed in Glass, <u>et al.</u>, 1972) have demonstrated that fixed effects ANCOVA F tests are robust with respect to violation of assumptions (2) and (3), providing that sample size is

Variable	Expe	Experimental Condition			
	Treatment I	Treatment II	Control		
CUE	79.88	76.44	72.31		
	( 7.91)	(9.38)	(11.45)		
PF	60.19	50.56	39.38		
	(12.59)	( 8.07)	(13.77)		
CUE-PF	84.50	75.25	57.13		
	(13.84)	(18.11)	(24.24)		
R-PF	9.44	6.31	7.00		
	(3.63)	(3.38)	( 4.65)		
Exam	73.63	71.88	70.25		
	(7.82)	( 6.14)	( 7.32)		

TABLE 13.--Means and Standard Deviations on the Dependent Variables and Covariate, by Experimental Condition.

TABLE 14.--Adjusted Means on the Dependent Variables, by Experimental Condition.

	Expe	Experimental Condition			
Variable	Treatment I	Treatment II	Control		
CUE	79.52	76.45	72.66		
PF	59.46	50.58	40.09		
CUE-PF	83.49	75.27	58.12		
R-PF	9.10	6.32	7.33		

moderately large (10 or more subjects per cell), and there is an equal number of observations per cell. Since both of these conditions were met in the present study, possible violation of assumptions (2) and (3) does not pose a threat to interpretation of the F statistics. Glass, et al., point out that little is presently known about the effect of violation of assumption (4) when the covariate is a random variable. However, they suggest that unless the departure from homogeneity of regression slopes is extreme, the effect on the F statistic is probably minimal. In order to determine whether this assumption was met, a test for homogeneity of regression was conducted for each dependent variable. The results of these tests, reported in Appendix I, indicated that for the variables PF, CUE-PF, R-PF, there was no evidence of departure from homogeneity of regression. Although the test for the variable CUE did indicate a lack of homogeneity, it was believed, upon further analysis, that this result could be attributed to a spurious negative regression weight for one group (Treatment II), and thus constituted a Type I error (see Appendix I for a more detailed discussion of the data relevant to this point). With respect to the first, and most crucial, of the ANCOVA model assumptions, it is believed that the experimental procedure included sufficient precautions to guarantee that this assumption was met. Although practical constraints required that subjects be assembled in groups for

training and testing, all instructions and materials were administered by means of individual booklets in selfinstructional format. Thus, it is believed that subject may be legitimately considered as the unit of analysis.

The multivariate analysis of covariance included one fixed independent factor (experimental condition) having three levels, with 16 subjects nested within each level, one covariate (Focal Problems Final Exam), and four dependent variables (CUE, PF, CUE-PF, R-PF). The ordering of the dependent variables for the conditional stepdown F tests was based on the following considerations. (1) Since performance on CUE (i.e., the detection and utilization of cues) is a prerequisite for the generation of problem formulation titles (PF performance), the variable CUE was ordered first and the variable PF second. Thus, for CUE the stepdown F test was the same as a univariate F test, while for PF the stepdown F provided a test of treatment effect on the generation of problem formulation titles with between-group differences on CUE partialled out. (2) Since the classification of cues with respect to problem formulations titles (CUE-PF performance) is a function of both cues obtained and problem formulation titles generated, CUE-PF was ordered third. Thus, the stepdown F ratio for CUE-PF provided a test of treatment effect on this variable with between-group differences on both CUE and PF partialled out. (3) Since statements of

functional relationships between problem formulations is dependent on the subject's prior processing of cues and generation of problem formulations (i.e., performance on CUE, PF and CUE-PF), this variable was ordered last. Thus, the fourth stepdown F ratio tests for treatment effect on R-PF with between-group differences on the other three variables partialled out.

The results of the MANCOVA are presented in Table 15. As shown in the table, the multivariate F test of equality of the vectors of adjusted group means was significant at p < .0052. Thus, for the set of four dependent

F tests	df	F	p
Multivariate F test	8, 82	3.0149	.0052
Stepdown F tests			
on CUE	2,44	1.9376	.1562
on PF	2,44	8.2725	.0010
on CUE-PF	2,44	.4261	.6559
on R-PF	2,44	1.8334	.1728

TABLE 15.--Multivariate Analysis of Covariance, on CUE, PF, CUE-PF, R-PF.

variables taken together, there was a significant main effect for treatment. The stepdown F tests yielded the following results: (1) no significant treatment effect on the variable CUE; (2) a significant treatment effect (p < .001) on the</pre> variable PF, conditioned on the variable CUE: (3) no significant treatment effect on the variable CUE-PF, conditioned on the variables CUE and PF; (4) no significant treatment effect on the variable R-PF, conditioned on the variables CUE, PF, CUE-PF.

In order to determine whether the nonsignificant stepdown F ratios occurred: (1) due to nonsignificant differences between group means, or (2) due to the fact that significant differences existed but had been partialled out in the calculation of the conditional stepdown F ratios, univariate F ratios were calculated on each dependent variable. The univariate ANCOVAS involved the same design model as the MANCOVA: namely, one fixed independent factor (experimental condition with subjects nested within levels of condition) and one covariate (Focal Problems Final Exam). The results of the ANCOVA on each dependent variable are presented in Table 16. As shown in the table, there was a significant treatment effect on the variables PF and CUE-PF (p < .0002, and p < .0020, respectively), but no significant treatment effect on the variables CUE and R-PF.

The results of the stepdown and univariate F tests indicate the following conclusions regarding each dependent variable.

 The differences among adjusted group means on the variable CUE are nonsignificant, as tested by a univariate F ratio.

Dependent Variable	Sources of Variation	df	MS	F	р
CUE	Group Subjects: Group Total	2 44 46	181.57 93.71	1.9376	.1562
PF	Group Subjects: Group Total	2 44 46	1446.76 131.36	11.0134	.0002
CUE-PF	Group Subjects: Group Total	2 44 46	2583.62 359.00	7.1966	.0020
R-PF	Group Subjects: Group Total	2 44 46	31.05 13.68	2.2698	.1154

TABLE 16.--Univariate Analyses of Covariance on CUE, PF, CUE-PF, R-PF.

2. The differences among adjusted group means on the variable PF are significant, as tested by a univariate F ratio or by a stepdown F ratio with PF conditioned on CUE. Thus, a significant treatment effect on PF is found not only when this variable is tested singly (by a univariate F ratio), but also when between-group variance on CUE is partialled out (by a stepdown F ratio).

3. The differences among adjusted group means on the variable CUE-PF are significant, as tested by a univariate F ratio. However, when CUE-PF is conditioned on CUE and PF (via a stepdown F ratio), differences among groups are not significant. Since significant betweengroup differences were found on PF but not on CUE, and since the within-group correlation of PF and CUE-PF was .75, the nonsignificant stepdown F for CUE-PF can be attributed to the partialling out of between-group differences on PF.

4. The differences among adjusted group means on R-PF were nonsignificant, whether measured by a univariate F ratio, or a stepdown F ratio.

The average within-group correlations between the covariate and the dependent variables were .15 for CUE, .26 for PF, .22 for CUE-PF, and .36 for R-PF. Of the four coefficients, only the one for R-PF was found to be significantly different from zero (F = 6.6840, df = 1, 44, p < .01). Given the nonsignificance of the three coefficients, and the low magnitude of the R-PF coefficient, we may conclude that the covariate was not effective in increasing the precision of the F tests. For the variables PF and CUE-PF significant univariate F tests were found in spite of the ineffectiveness of the covariate. For the variables CUE and R-PF, it is possible that significant F tests would have been obtained if a more powerful covariate had been available.

Having found a significant univariate treatment effect on the variables PF and CUE-PF, the Scheffé <u>post hoc</u> confidence interval procedure was used in order to test for significant differences on these variables between each pair of experimental conditions. Although the more

powerful Tukey <u>post hoc</u> procedure would generally be used when simple pair-wise comparisons are desired and there is the same number of subjects per level of the independent variable, this procedure is not considered applicable following analysis of <u>covariance</u> because the estimates of the regression intercepts do not in general meet the requirement of equal variances and covariances (Scheffé, 1959, cited in Glendening, 1973).

The results of the Scheffé <u>post hoc</u> procedure are presented in Table 17. Similar results were found for both dependent variables: (1) the difference between the two treatment groups is nonsignificant; (2) the difference between Treatment I and the control group is significant at the .001 or .005 level; (3) the difference between Treatment II and the control group is significant at the .05 level.

The results on the preceding analysis will now be discussed with respect to each of the experimental hypotheses.

Hypothesis 1:

The average performance of second-year medical students who have received problem formulation training (Treatment I and Treatment II) will be superior to that of students who have not received training (control group), as measured by four dependent variables: (1) CUE score, (2) PF score, (3) CUE-PF score, and (4) R-PF score.

Variable	Comparison	Confidence Interval	(1-∝)%
PF	T1-T2 8.8767	±10.3227	95%
	T1-C 19.3699*	±16.6644	99.9%
	T2-C 10.4932*	±10.3187	95%
CUE-PF	T1-T2 8.2107	±17.0648	95%
	T1-C 25.3706*	±23.6577	99.5%
	T2-C 17.1599*	±17.0583	95%

TABLE 17.--Scheffé post hoc Comparisons on PF and CUE-PF.

\*Significant group differences at the .05 level or better.

NOTE: Comparisons are made on adjusted group means. T1 = Treatment I; T2 = Treatment II; C = Control.

The results of the analysis supported this hypothesis with respect to the variables PF and CUE-PF, but not with respect to the other two variables.

On the variable CUE there was no significant difference between the treatment group means and the control group mean. If the means on CUE are expressed as a percentage of the maximum possible score on this variable, it is found that the average performance under all three conditions was high (77.2% for Treatment I, 74.2% for Treatment II, 70.5% for the control group). Thus, we may conclude that the subjects had already attained, prior to the experiment, a high level of ability with respect to detecting cues, and utilizing them to generate at least one problem formulation.

On the variable PF the treatment group means each differed significantly from the control group mean. Since the treatment and control subjects did not differ on the variable CUE, the significant differences on PF cannot be attributed to a failure on the part of the control subjects to acquire sufficient cues to generate problem formulations. Thus, we may conclude that the effect of the training was to improve the subject's skill in making use of the cues he obtained in order to generate a thorough and appropriate set of initial problem formulations.

On the variable CUE-PF the treatment group means each differed significantly from the control group mean. Thus, the training was also effective in improving subjects' performance on the task of classifying cues with respect to the problem formulation categories of major importance for the case. However, the results of the stepdown F test on CUE-PF indicates that between-group differences on this variable can be attributed to the between-group differences that occurred on PF. Thus, we may conclude that although training significantly improved the subject's performance

on CUE-PF, this effect was a function of improvement in the thoroughness and appropriateness of the problem formulations he generated.

On the variable R-PF there was no significant difference between treatment group means and the control group mean. Thus, we may conclude that the training had no effect on the subject's ability to hypothesize possible functional relationships among the problem formulations he generated.

#### Hypothesis 2:

The average performance of second-year medical students who have received problem formulation training involving outcome and process feedback (Treatment II) will be superior to that of students who have received problem formulation training involving only outcome feedback (Treatment I), as measured by the four dependent variables: (1) CUE score, (2) PF score, (3) CUE-PF score, and (4) R-PF score.

There were no significant differences between the two treatment groups on any of the variables. Thus, the second experimental hypothesis was not supported. Moreover, the direction of observed differences indicated a trend in the opposite direction than that hypothesized: namely, the means for the "outcome feedback only" condition were consistently higher than the means of the "outcome plus process feedback" condition. Except for CUE, the Treatment I group means did not closely approach the maximum possible score on each variable (the PF mean was 50.8% of the maximum, the CUE-PF mean 40.1% of the maximum, and the R-PF mean 37.9% of the maximum). Thus, the lack of significant differences between treatment groups cannot be attributed to a ceiling effect. As indicated in Table 14, differences between the treatment conditions were quite sizable on the variables PF and CUE-PF. It is possible that these differences would have proved to be significant if the covariate had been more powerful. Thus, we may conclude that while neither treatment was found to be more effective than the other, the results suggest a possible superiority of Treatment I over Treatment II.

Further interpretation of the results of the hypothesis tests will be undertaken in the second section of this chapter.

### Relationships Among Dependent Variables

As described in Chapter III (p. 94), each dependent variable scoring key was designed to measure a distinct component of the subject's performance on the posttest task. task. It was assumed that the measures would show moderate positive intercorrelations, but that no correlation would be so high as to indicate that performance on one variable can be fully predicted by performance on any other(s). In particular, it was argued that the ability to classify cues with respect to problem formulation categories (CUE-PF performance) would not be a simple linear function of performance on the two single-dimension variables (CUE

and PF). The keys were constructed so that even though two subjects had identical CUE and PF scores they could differ in performance on CUE-PF (e.g., one could show greater flexibility in associating appropriate cues with the <u>multiple</u> problem formulations to which these cues were potentially relevant). The results of the stepdown F tests indicated that, at least so far as between-group differences are concerned, performance on CUE-PF could be predicted by performance on PF. In order to determine if this were also true at the within-group level, a within-group multiple linear regression of CUE-PF on PF and and CUE was carried out.

The intercorrelations among the dependent variables, as well as the results of the multiple regression analysis, are presented on Table 18. All correlations, except for that between PF and CUE-PF, were as anticipated: positive and moderate. The results of the regression analysis indicated that a very sizable portion of the variance on CUE-PF could be accounted for by PF and CUE ( $R^2 = .62$ ). The estimates of variance accounted for by step-wise addition of PF and then CUE to the equation indicated that PF alone accounted for 56% of the variance, while the addition of CUE accounted for only an additional 6% of the variance. Since the multiple correlation coefficient (R = .79) is probably as high as could be attained given the reliability of the measures, we are led to conclude that within-group

	Average Within Depe	Average Within-Group Correlations Among Dependent Variables				
	CUE	PF	CUE-PF	R-PF		
CUE	1.00					
PF	.34	1.00				
CUE-PF	.48	.75	1.00			
R-PF	.27	.41	. 39	1.00		
	Multiple Regression of CUE-PF on CUE and PF (Within-group)					
	Regression Weight		Partial Correlation			
PF	1.09		.67			
CUE	.50		<b>.2</b> 5			
	Multiple R		Multiple R <sup>2</sup>			
	• 79		.62			
	% of Varianc Step-w	e Accour ise Addi	nted for by			
PF		56%				
CUE		68				
Total		628				

TABLE 18.--Relationships Among Dependent Variables.

performance on CUE-PF is a linear function of performance on PF and CUE. In sum, it appears that once a subject has obtained the cues presented during the workup, and generated a set of problem formulation titles, the task of determining which cues are of relevance to each title does not pose any further difficulty.

#### Supplemental Analyses

This section will present a number of supplemental analyses that were conducted to aid in interpretation of the experimental outcomes described in the preceding section.

### Results of the Additional Posttest Tasks

As explained in Chapter III (p. 143), two additional tasks were administered at the posttest session in order to determine whether failure in the processes of cue detection, encoding and retrieval may have inhibited performance on the basic posttest task as measured by the four dependent variables. For reasons discussed previously, the additional tasks pertained only to film 8.

It will be recalled from Chapter III that the CUE score was a weighted sum of points obtained for each cue a subject listed under at least one problem formulation title. Thus, high performance on CUE in itself provides evidence that failure in cue acquisition was not responsible for low performance on the PF variable. The additional posttest tasks were administered in order to aid in interpreting the experimental outcomes in the event that there were significant between-group differences on CUE, and that these differences were predictive of differences on PF. Since performance on CUE was uniformly high across

all three experimental conditions, the data from the additional posttest tasks merely corroborates the conclusions reached in section one of this chapter.

The subject's performance on the Recognition of Cues task was summarized in terms of the number of each type of item he checked: (1) number of cues (out of 32); (2) number of consistent distractors (out of 16); (3) number of contradictory distractors (out of 8); (4) number of inconsistent distractors (out of 8). In some instances a subject failed to check a cue on the recognition task even though he had listed it under one of his problem formulations in carrying out the basic posttest task. Since the primary purpose of this analysis was to determine the number of cues the subject had obtained from the film and could have potentially used in generating problem formulations, an additional variable was also calculated: number of cues obtained (i.e., number of cues checked on recognition task + number of cues used on basic posttest task but not checked on recognition task). Group results on each of these measures are presented in Table 19. As shown in this table an average of 30 (out of 32) cues was obtained under each experimental condition. Moreover, no subject obtained less than 27 cues. Clearly, detection, encoding and retrieval of cues presented no obstacle to carrying out the basic posttest task, and in no way contributed to between-group differences on PF. A

Variable	Treatment I	Treatment II	Control
No. cues obtained	30.1	30.1	29.4
	(28-31)	(28-31)	(27-31)
No. cues checked	29.5	29.9	28.5
	(27-31)	(28-31)	(25-31)
No. distractors checked consistent	1.1	2.0	1.3
	(1-3)	(0-6)	(0-3)
contradictory	0.1	0.2	0.3
	(0-1)	(0-1)	(0-2)
inconsistent	0	0	0

TABLE 19.--Results of the Recognition of Cues Task, by Experimental Condition.

tabulation of the cues that were not obtained (i.e., a total of 97 omissions across all 48 subjects' responses), revealed that in 80 instances (82.5%) the cue had weight of one in the CUE scoring key. Twelve of the 97 omissions (12.3%) involved cues with a weight of two; and only 5 omissions (5.2%) involved cues with a weight of three. Thus, only very rarely did a subject fail to obtain cues that were of major importance for the generation of appropriate problem formulations.

Examination of the cues that had been listed by the subject on the basic posttest task, but not checked on the Recognition of Cues task (i.e., cues obtained - cues checked), indicated that in 23 instances (out of a total of 25 such instances across all 48 subjects), the cue forgotten had a weight of one and/or was very closely related to other cues that were checked. Of the 20 subjects who forgot cues, only 4 forgot more than one cue. Thus, not only was the forgetting of cues between the two tasks very minimal, when it did occur it usually involved single items that were of minor importance and/or highly redundant with other cues that were recalled.

Examination of the distractors checked on the recognition task indicated that errors of commission were as infrequent as errors of omission. The extremely low frequency with which the contradictory distractors were checked indicated that if a subject detected a cue while viewing the film, he nearly always encoded it correctly. The results for the other two types of distractors indicated that if a subject "recalled" pieces of data that were not in fact presented in the film, these were always data which were consistent with the cues that were presented. Thus, to a very limited degree subjects showed a tendency to "supplement" the nominal stimulus by generating, as "cues," items that presumably are closely associated with the actual cues in long-term memory.

In the second additional posttest task, the subject was provided with a list of the cues presented in the film and asked to make any additions he wished to his original response sheets. The purpose of this task was to determine

if performance on the basic posttest task would have been higher if the process of generating problem formulations had not been dependent on the subject's detection, encoding and retrieval of cues. For each subject a new set of PF, CUE and CUE-PF scores were calculated on the basis of his initial responses plus his additions to his response sheets. As discussed in Chapter III, additions to response sheets could occur for two reasons: (1) because the list of cues provided the subject with data which he had failed to obtain while viewing the film, and (2) because the list provided the subject with a second exposure to cues he had originally obtained, but failed to utilize in generating problem formulations. Since the present analysis was concerned only with the first factor listed above, the following criteria were used in determining which additions a subject made would be included in the calculation of his new dependent variable scores. (1) Problem formulation titles were counted as additions providing that at least one of the cues listed under this title had not been previously obtained (i.e., had not been included in the subject's basic posttest responses, and/or had not been checked on the recognition checklist). (2) Cues were counted as additions only if not previously obtained (as indicated by the subjects basic posttest and/or checklist responses).

The results of the additions task are presented in Table 20. There was very little change in the group means

	Treatment I	Treatment II	Control
Group means			
CUE	45.88	43.88	40.25
CUE (A)	47.32	44.69	42.00
PF	34.44	29.88	23.19
PF (A)	34.94	30.82	23.82
CUE-PF	50.44	42.44	33.19
CUE-PF (A)	51.25	43.88	35.07
Mean increment for subjects whose scores changed			
CUE	3.29	2.17	2.80
(n)	(7)	(6)	(10)
PF	2.67	7.50	5.00
(n)	(3)	(2)	(2)
CUE-PF	2.60	4.60	6.00
(n)	(5)	(5)	(5)

TABLE 20.--Results of the Additions to Response Sheets Task, by Experimental Condition.

NOTE: Group means are reported on each variable for the subjects' original scores on the basic posttest task, and for the subjects' composite scores (including additions), designated by (A) following the variable label.

on PF, due to the fact that only two or three subjects in each group generated any additional problem formulations. A sizable number of subjects in each group improved their CUE and CUE-PF scores, but not by a very large amount on the average. On all variables the increments in the group mean was fairly constant across experimental conditions. Thus, the between-group experimental outcomes found on the basic posttest task were not altered by providing the subject with the cues he initially failed to obtain. This finding corroborates the conclusion drawn from the hypothesis tests reported earlier: namely, treatment-control differences in the generation of problem formulations cannot be attributed to differences in cue acquisition.

# Treatment-Control Differences in Problem Formulations

Several supplemental analyses were undertaken in order to determine more precisely the nature of the significant treatment-control differences that were found on the variable PF.

The first analysis was concerned with the structural properties of the "problem spaces" generated by subjects under each experimental condition. It will be recalled from Chapter IV, that four features were found to be characteristic of the sets of problem formulations generated by the experienced physicians: (1) hierarchial organization; (2) competing formulations, (3) multiple subspaces, and (4) functional relationships. In addition, it was found that the size of the physician's set of formulations could be measured in terms of (1) the number of problem formulations it contained, and (2) the number of subspaces it contained. Analysis of the student data in terms of these six variables yielded the results presented in Table 21.

Variable	Treatment I	Treatment II	Control
Structural Features <sup>a</sup>			
Film 7			
Hierarchical organization Competing formulations Multiple subspaces Functional relationships	12 16 16 15	15 16 16 13	9 11 15 11
Film 8			
Hierarchical organization Competing formulations Multiple subspaces Functional relationships	16 15 16 14	14 12 16 8	9 8 16 10
Problem Space Size <sup>b</sup>			
Film 7			
Number of problem formulations	6.9 (4-11)	5.5 (3-7)	4.2 (1-8)
Number of subspaces	4.0 (3-6)	3.7 (3-5)	3.1 (1-5)
Film 8			
Number of problem formulations	8.4 (5-13)	6.8 (3-10)	5.1 (3-9)
Number of subspaces	4.4 (3-5)	<b>4.</b> 2 (3-5)	3.3 (2-5)

TABLE 21.--Analysis of the Structure of the Students' Sets of Problem Formulations, by Experimental Condition.

<sup>a</sup>Table entries are the number of subjects whose set of problem formulations exhibited each feature.

<sup>b</sup>Table entries are the mean and range on each variable.

An examination of the data on structural features indicated that the control group tended to differ from the treatment groups with respect to two features: (1) hierarchical organization, and (2) competing formulations. The fact that fewer control subjects generated hierarchically organized sets of problem formulations can in part be attributed to the fact that they generated fewer problem formulations, and therefore had less need to use hierarchical organization as a means of increasing working memory storage capacity.

The relative infrequency of competing formulations among control subjects is a more critical manner. As indicated in Chapter IV, competing formulations was the one feature that was found to characterize all physicians' performance on every film. Thus, a major difference between the control subjects' responses and those of the trained subjects (especially under Treatment I) was that many of the former (approximately one-half to one-third, depending on the case) failed to generate formulations having the feature that was uniformly characteristic of the experienced physicians, while nearly all of the latter did so. This finding would suggest that one effect of the training procedure, at least under "outcome feedback only" condition, was to improve the subject's skill in generating competing formulations.

In order to determine whether the observed treatmentcontrol differences on the measures of problem space size were statistically significant, a multivariate analysis of covariance was conducted. For the purpose of this analysis, the subject's scores on the two variables were summed across the two tasks. The results of the analysis, reported in Table 30, Appendix H, revealed: (1) a significant multivariate main effect (F = 5.0314, p < .0011), (2) a significant main effect on number of subspaces (univariate F = 6.2399, p < .0042), (3) a significant main effect on number of problem formulations conditioned on number of subspaces (stepdown F = 4.0060 p < .0254). Scheffé post hoc comparisons, were then carried out on the adjusted group means for each pair of experimental conditions. The results of these comparisons, reported in Table 31, Appendix H, were as (1) the Treatment I mean was significantly higher follows: than the control mean on both number of problem formulations and number of subspaces (p < .001 and p < .05, respectively), (2) the difference between the Treatment II and the control group was not significant on either variable; (3) the difference between the treatment groups was not significant on either variable.

On the basis of this analysis we may conclude that the treatment-control difference on the variable PF can be attributed in part to the fact that the trained subjects (at least under Treatment I) generated a larger number of

both problem formulations and subspaces than the control subjects. The larger number of subspaces generated by the Treatment I subjects indicates that one effect of the "outcome feedback" training was to increase the scope (or horizontal dimension) of the student's problem spaces. Moreover, the significant stepdown F ratio for number of problem formulations conditioned on number of subspaces, coupled with the results of the <u>post hoc</u> comparisons (indicating that the Treatment I-control difference in number of problem formulations cannot be accounted for by the between-group difference in number of subspaces) suggests that the "outcome feedback" training also led to an increase of problem space size on the vertical dimension of hierarchical elaboration within subspaces.

It is also of interest to consider the data on subspaces with respect to the parameter of memory organization that has been proposed by Mandler (1967). The range of subspaces (per film) generated under both treatment conditions coincides very closely with Mandler's proposition that human information-processors typically organize and store items in terms of  $(5 \pm 2)$  categories. Under the control condition, the number of subspaces generated never exceeded the upper limit of Mandler's parameter, but, in the case of five subjects on each task, did fall below the lower limit.

The data on number of subspaces provide a quantitative measure of the scope of subjects' problem spaces, but do not indicate whether the  $(5 \pm 2)$  subspaces typically generated were the most appropriate ones for the case. In order to address this question, a second type of analysis was undertaken. In this analysis, the subjects' problem formulation responses were tabulated in terms of eight major problem formulation categories (plus two specific formulations) that were found to characterize the performance of all of the experienced physicians on the two posttest films.<sup>1</sup> Each of the eight categories (listed on Table 22) represents a different problem subspace. The table indicates the number of subjects, under each condition, who generated at least one problem formulation within each of the eight categories. It also indicates the number of subjects who generated two specific formulations of particular importance for film 8, and the number of subjects generating at least one formulation within various combinations of categories.

For film 7, the table indicates that all subjects generated problem formulations pertaining to the patient's recent (acute) symptoms of respiratory infection. In this case, however, the problematic aspect is to recognize that

<sup>&</sup>lt;sup>1</sup>These categories, it may be noted, and were also the categories used in the CUE-PF scoring key.

Problem Formulation Categories of Major			
Importance	Treatment I	Treatment II	Control
Film 7			
<ol> <li>Acute respiratory infection</li> <li>Chronic respiratory problem</li> <li>Cancer</li> </ol>	16 15	16 14	16
Categories 1 plus 2 or 3	16	14	14
All three categories	15	12	6
Film 8			
<ol> <li>Pregnancy         <ul> <li>Ia. Ectopic pregnancy</li> <li>Psychological problem</li> <li>Gastrointestinal probl 3a. Appendicitis</li> <li>Diabetes mellitus</li> <li>Genito-urinary infecti</li> </ul> </li> </ol>	16 8 14 em 13 6 15 on 12	16 8 16 13 7 13 7	16 5 16 8 4 9 3
At least four categories	15	12	7
All five categories	7	5	3

TABLE 22.--Number of Subjects Generating at Least one Problem Formulation in Various Categories, by Experimental Condition.

the patient's acute problem is superimposed on a more serious and long-term problem (chronic respiratory disease, or cancer). All but two control subjects generated formulations in one of these latter categories thereby indicating their recognition of the dual level nature of the case. However, only six control subjects as compared to 15 subjects under Treatment I and 12 under Treatment II, generated formulations in

.

both of these categories. Thus, the major difference between the treatment and control group was the latter's failure to generate multiple competing formulations pertaining to the patient's underlying problem.

For film 8, there was a considerably larger number of major problem formulation categories to be considered. Since the patient is a single student nurse who thinks she may be pregnant, and has broken up with her boyfriend, it is not surprising that nearly all subjects generated the formulations of pregnancy and psychological problem. The main difference between the treatment and control groups was their skill in generating multiple formulations to account for the patient's other complaints (severe abdominal pain and vomiting for one day's duration; increased appetite, thirst and urination for several weeks; a recurrent vaginal itch). Although all of these symptoms might be attributed to pregnancy and/or a psychological problem, they strongly suggested (to all experienced physicians) that three other categories should be considered: namely, gastrointestinal problem, diabetes mellitus, and genito-urinary tract infection. Only seven control subjects generated problem formulations pertaining to at least four out of the five categories, whereas 15 of the Treatment I subjects and 12 of the Treatment II subjects did so. The patient in film 8 could easily have multiple, concurrent problems (e.g., pregnancy, psychological problems,

a G.U. infection and diabetes), but it is also necessary to consider competing explanations for some of her symptoms (e.g., diabetes versus G.I. disorder versus ectopic pregnancy). Thus, the film 8 data suggest that the control subjects' performance was inferior to that of the trained subjects (at least under Treatment I) both with respect to the scope of appropriate subspaces considered, and with respect to the number of competing formulations generated.

It may be noted that one inadequacy of the students' performance under all three conditions was the relatively small number of subjects (4-8) who generated the two specific formulations of ectopic pregnancy and appendicitis. Nearly all physicians indicated that these formulations should be considered early in the workup of the film 8 case due to their immediate life-threatening implications.

To summarize: the results of the supplemental analyses corroborate the conclusions drawn from the hypothesis tests on the variable PF, namely, that the trained subjects (especially under Treatment I) generated more thorough and appropriate sets of problem formulations. Treatment I group performance was found to be superior not only on a quantitative dimension (i.e., number of problem formulations and subspaces generated), but also on a qualitative dimension (i.e., number of subspaces of major importance in which at least one formulation was generated).

# Comparison of the Treatment Conditions

The hypothesis tests reported in the first section of this chapter indicated that there were no significant differences between the two treatment groups. However, it was noted that on every variable examined, in both the major and supplemental analyses, the direction of the difference between the groups was in favor of Treatment I. It is possible that if a more powerful covariate had been employed the observed differences on many of these variables would have been statistically significant. Alternatively, if one were to accept a higher probability of a Type I error (namely,  $\alpha = .10$ ), most of the differences between treatment groups would be found significant. Thus, there was some evidence to suggest that, if either training procedure is to be preferred, it is the "outcome feedback only" procedure rather than the "outcome plus process feedback" procedure. This final section of Chapter V will first present the results of some supplemental analyses regarding the treatment groups, and second address the question as to why the process feedback was so ineffective.

PF scoring keys were prepared for each of the six training films, and all subjects' responses scored by the experimenter. Table 23 presents the treatment group means and standard deviations on PF for each of the films. As indicated in the table, the group means are almost identical on the first film, but begin to diverge on the second film,
	Treat	ment	
Training Film	Ī	II	
1	20.25 (4.24)	20.44 (5.29)	
2	36.44 (7.16)	33.31 (7.13)	
3	18.06 (3.36)	17.19 (3.25)	
4	33.56 (5.81)	29.50 (8.34)	
5	21.50 (6.20)	<b>19.56</b> (5.66)	
6 <sup>a</sup>	30.25 (6.29)	25.19 (4.26)	

TABLE 23.--Means and Standard Deviations of Treatment Group PF Scores on the Training Films (1-6).

<sup>a</sup>The treatment group means on Film 6 are significantly different at p < .01, F = 7.10732, with 1 and 30 degrees of freedom.

and on the last film are significantly different at the .01 level. Although there is no clear trend across the training tasks, it is possible that with a longer period of training the cummulative effects of the treatments would have led to a significant contrast between the groups on the posttest, and, thus, have provided clear evidence of the superiority of Treatment I on the variable PF.

A second supplemental analysis was based on the responses to the questionnaire, administered at the end of

the posttest session in order to determine the students' opinions of the training procedures and materials. The subject's response to each item in the questionnaire (section 1) was scored on a five-point scale: +2 = strongly agree; +1 = agree; 0 = no opinion; -1 = disagree; -2 = strongly disagree. Table 24 reports the group means and standard deviations for each item. In addition, each subject's score on three summary variables was calculated. These variables were as follows:

EV FILM: the subject's evaluation of the six
 filmed interviews (i.e., the mean of his responses to items
 3-9, with the sign reversed for item 5).

2. EV FB: the subject's evaluation of the feedback materials (i.e., the mean of his responses to items 10, 11, 12 and 13, with the sign reversed for item 11).

3. EV GEN: the subject's evaluation of the overall effectiveness of the training materials and procedures (i.e., the mean of his responses to items 12, 17 and 20).

The group means and standard deviations on these variables are reported in Table 25. In order to test the significance of the differences in the group means, a oneway fixed effects analysis of variance was performed on each variable. The results of these analyses are found in Table 26.

		Trea	tment
	Item	I	II
1.	The instructions were generally clear and easy to follow.	1.438 (.512)	1.125 ( .957)
2.	The instructional sessions were too long.	375 (1.088)	188 (1.109)
3.	The actors who played the role of the patients in the films were very convincing.	1.438 ( .629)	1.438 (.512)
4.	The films provided a realistic simulation of the early part of the clinical workup.	1.438 ( .814)	1.188 ( .544)
5.	The dialogue in the films was some- times difficult to follow.	-1.125 ( .342)	-1.063 ( .680)
6.	The physicians in the films did a good job of interviewing the patients.	.563 (.727)	.125 (1.088)
7.	I enjoyed watching the films.	1.438 ( .629)	1.250 (.577)
8.	As I watched the films, I was able to put myself into the role of the doctor	.875 .(.885)	.438 (1.094)
9.	The films presented a good selection of medical cases.	1.250 (.775)	.688 (.704)
10.	The feedback materials were well organized and easy to follow.	1.375 (.500)	.875 (.806)
11.	The feedback materials were sometimes overly redundant.	313 ( .873)	.563 (.892)
12.	The opportunity to compare my problem formulations to those of experienced physicians helped to improve my skill in generating initial problem formu lations.	1.000 (.365)	.938 (1.063)
13.	I found the feedback materials interesting.	1.438 ( .512)	1.000 ( .632)
14.	Treatment I: The second viewing of the film helped me to consolidate my understanding of the case.	500 (1.155)	

!

TABLE 24.--Mean and Standard Deviation of Treatment Group Responses to Questionnaire Items (Section 1).ª TABLE 24.--Continued.

	Treat	ment
Item	I	II
Treatment II: The second version of the films, which portrays the phy- sician "thinking aloud," provided me with an understanding of the process by which experienced physicians generate initial problem formulations.		.625 (1.025)
15. Treatment I: The second viewing of the film was not worthwhile.	.625 (1.147)	
Treatment II: The "think aloud" segments in the second version of the films tended to disrupt my own thinking process.		750 ( .931)
16. The self-evaluation checklists helped me to evaluation my own performance as compared to that of the experi- enced physician.	1.000 (1.095)	.750 (1.125)
17. My ability to generate a set of initial problem formulations has improved as a result of utilizing this instructional package.	1.188 ( .655)	.563 ( .814)
18. This instructional package is not appropriate for second-year medical students.	-1.438 ( .512)	-1.500 ( .516)
<pre>18a. It would be more appropriate for first-year students.</pre>	-1.000 ( .731)	813 ( .911)
18b. It would be more appropriate for third-year students.	438 ( .814)	875 ( .885)
19. For some of the cases, I didn't have sufficient medical knowledge to be able to generate appropriate problem formulation.	.125 (1.147)	.125 (1.025)
20. If a library of films like these, with accompanying feedback materials, was available to medical students, I would make use of it.	1.125 ( .619)	1.000 ( .632)
21. It would be more interesting to use the films and feedback materials in a group discussion setting (e.g., focal problems class) than in an individual self-instructional format.	.375 (1.088)	.625 (1.147)

<sup>a</sup>Ss responded to each item on a five-point scale (+2= strongly agree; +1=agree; 0=no opinion; -1=disagree; -2= strongly disagree).

	Trea	tment	
Score <sup>a</sup>	I	II	
EV FILM	1.1607 (.397)	.884 (.428)	
EV FB	1.031 (.352)	.563 (.452)	
EV GEN	1.104 ( .359)	.833 (.632)	

TABLE 25.--Means and Standard Deviations of Treatment Group Scores on Questionnaire (Section 1).

<sup>a</sup>Range of scale is from +2 (highly positive) to -2 (highly negative).

TABLE 26.--Analyses of Variance on the Questionnaire Scores: EV FILM, EV FB, EV GEN.

Score	Sources of Variation	df	MS	F	p
EV FILM	Between groups Within groups	1 <u>30</u> 31	.6129 .1702	3.6011	.067
EV FB	Between groups Within groups	1 <u>30</u> 31	1.7578 .1641	10.7143	.003
EV GEN	Between groups Within groups	1 <u>30</u> 31	.5868 .2646	2.2179	.147

A review of the data in these tables suggests the following conclusions regarding the students' opinions of the training materials and procedure.

First of all, it may be noted that, with one exception, both groups evaluated the films, feedback materials and training procedure as a whole in a positive manner (as indicated by mean scores close to 1.0 on each variable in Table 25). The one exception was the Treatment II mean on EV FB which was half-way between the positive and neutral points on the scale. Thus, we may conclude that the students in both groups reported a generally positive opinion of the materials and procedure.

There were, however, some differences between the groups with respect to their opinions. As indicated in Table 26, the Treatment I mean on EV FB was significantly higher than the Treatment II mean (p < .003). Thus, on the factor which differentiated the two groups--type of feedback-the Treatment I group's opinion was more favorable than that of the Treatment II group. Although the groups did not differ significantly on the other two variables (EV FILM, EV GEN), it should be noted that Treatment I had slightly higher means, and slightly lower standard deviations, on these variables than Treatment II, indicating that at the level of the individual subject there were more persons reporting relatively unfavorable opinions under Treatment II. This observation is also borne out by an examination of the means and standard deviations for single items (see Table 24).

The responses to the final six items in section 1 of the questionnaire provide data on two further points. First, although some subjects in both groups felt that they didn't have sufficient medical knowledge to generate appropriate problem formulations for some of the cases (item 19), nearly all subjects in both groups indicated that they found the training package to be appropriate for secondyear students (item 18). Secondly, the students in both groups indicated generally favorable attitudes toward eventual application of the training materials in a selfinstructional and/or group discussion setting (items 20 and 21).

To summarize: Analysis of the responses to section 1 of the questionnaire yielded results that closely paralleled those obtained from the analysis of the posttest data. Differences between treatment groups were largely nonsignificant, but there was some evidence to suggest that the "outcome feedback only" training elicited more favorable student opinions.

We will now address the question of why providing the subject with process feedback, in addition to outcome feedback, did not lead to superior performance on the part of the Treatment II subjects, as was originally anticipated. First of all, it is necessary to consider the possibility that "outcome plus process feedback" <u>is</u> superior to "outcome feedback only," but that its effectiveness was not detected

due to some failure in the experimental procedure. Two potential sources of internal invalidity will be considered.

1. Failure of random assignment to yield equivalent groups. Although one can never be sure, in any single experiment, that the treatment groups are actually (as opposed to randomly) equivalent, there is little basis for hypothesizing this factor as an explanation of the ineffectiveness of process feedback. The groups had highly similar scores on the covariate. Moreover, on a second variable of potential relevance to the dependent measures: namely, amount of clinical experience prior to participating in the experiment, any difference that did exist was in favor of the Treatment II group (see Table 27).

2. Extra-session or intra-session "history" as a confounding variable. "History" is the term applied by Campbell and Stanley (1969) to events which occur concurrently with, and are confounded with, the administration of treatment. In the present study, there are two ways in which the experimental design could have failed to control for the confounding effects of history. The first is <u>extra-session</u> history: principally, the possibility that subjects in one treatment group pursued an interest in the training cases outside of the experimental sessions to a greater degree than subjects in the other group. The second section of the questionnaire was designed to ascertain if this occurred. The subject was asked to indicate for each case: (a) whether he discussed it with other students

	Treatm	ent
Variable	I	II
ursuit of interest in raining cases outside the sessions		
No. of cases:		
a) discussed with students		
Mean Range	3.8 (0-6)	3.8 (0-6)
b) discussed with faculty		
Mean Range	0.4 (0-3)	0.1 (0-2)
c) looked up references		
Mean Range	1.4 (0-4)	1.4 (0-4)
d) any of (a), (b), (c)		
Mean Range	4.1 (0-6)	4.4 (0-6)
linical experience prior o participation in experiment		
(40-hour weeks)		
0-12 weeks	n = 13	n = 10
13-52 weeks	n = 2	n = 1
53 or more weeks	n = l	n = 5

TABLE 27.--Responses to Sections Two and Four of the Questionnaire.

in his group; (b) whether he discussed it with faculty members; (c) whether he looked up reference materials of relevance to the case. The students responses to these queries are reported in Table 27. As indicated in the table, there was no evidence of between-group differences with respect to extra-session pursuit of interest in the training cases.<sup>1</sup>

It is always possible, of course, that the groups' extra-session history systematically differed in some other way, but this is implausible given the homogeneity of the students' curricular activities.

A second possibility is that of between-group differences in <u>intra-session</u> history, due to the fact that training was administered in group sessions. This weakness in the experimental design was recognized at the outset, but could not be avoided because of practical constraints. It is believed that the use of individual self-instructional booklets to administer the training provided sufficient control against this source of internal invalidity. No events occurred during the sessions to suggest that there systematic between-group differences in intra-session history. Nevertheless, this possibility cannot be ruled out of consideration.

<sup>&</sup>lt;sup>1</sup>It may be noted that the data in Table 27 also bear on the issue, previously raised, of the students' attitude toward the training materials, with the level of the group means tending to indicate a high level of interest in the training cases on the part of students in both groups.

We will now consider an alternative interpretation for the outcome of the experiment: namely, that the process feedback was in fact ineffective. Assuming that failures in experimental method did not occur, the results indicate that providing the subject with process feedback, in addition to outcome feedback, clearly did not have a positive effect on the development of his problem formulation skills, and may even have had a negative effect, as compared to outcome feedback only. The experimenter's interpretation of this phenomenon rests on two hypotheses. The first is that, having been given the outcome feedback, the Treatment I subjects had no difficulty in inferring what the physicians' reasoning process must have been in order to generate the formulations listed on the outcome feedback Thus, they were able to provide themselves with sheets. self-generated process feedback, and thereby received, in essence, the same "treatment" as the other group. This factor could explain equivalence of posttest performance by the two groups. However, a second hypothesis is needed in order to account for the evidence suggesting a possible superiority of outcome feedback alone. It would appear that the Treatment II condition may have provided the subject with too much feedback, and thereby led to a diminishing of interest in the task. Two observations support this hypothesis. First, the experimenter noted that during the presentation of the process feedback films some of the

Treatment II subjects did not appear to be actively attending to the film. Second, the questionnaire item on which the largest difference was found between the groups was number 11 (see Table 24), on which the majority of Treatment II subjects agreed that "the feedback materials were sometimes overly redundant" while the majority of Treatment I subjects disagreed. In response to another item (14), the majority of Treatment II subjects indicated that the supplemented films did convey "an understanding of the process by which experienced physicians generate initial problem formulations." However, the real issue seems to be whether the films were necessary to this effect, or whether <u>self</u>-<u>generated</u> process feedback, as apparently occurred under the Treatment I condition, is not more effective from both a cognitive and a motivational standpoint.

A third interpretation to be considered is that process feedback could <u>potentially</u> be effective but was not in this experiment due to inadequacies in the data that were obtained from the experienced physicians. In this study, as in other recent investigations of medical problem solving, introspection was the technique used to obtain data on cognitive processes. As reported in Chapter IV, the strategies for generating problem formulations that were found to be employed by a sizable number of physicians (and thus were included, along with other process-related commentary, in the "think aloud" segments of the films)

were neither large in number nor very complex, a factor which may help to account for the Treatment I subjects' apparent ability to generate their own process feedback. It is always possible, however, that the physician does employ a variety of complex strategies to generate initial problem formulations, but that these strategies have become so routinized as to be inaccessible by means of introspective self-report. If this were the case, and if it were possible to identify these strategies (by using, for example, an approach of the Bruner, Goodnow and Austin (1956) type<sup>1</sup>), then it might be found that the provision of process feedback <u>would</u> significantly increase the effectiveness of the training model.

<sup>&</sup>lt;sup>1</sup>The Bruner, Goodnow and Austin (1956) approach attempts "to <u>externalize</u> for <u>observation</u> as many of the decisions as could possibly be brought into the open in the hope that regularities in these decisions might provide the basis for making inferences about the processes involved . . . " (p. 54) (my italics).

### CHAPTER VI

### CONCLUSIONS AND IMPLICATIONS

This chapter will summarize the major conclusions of this research, and will indicate some implications of these conclusions for future research and instructional development.

## Problem Formulation Outcomes and Processes in the Experienced Physician

Conclusions drawn from the analysis of the physician data are necessarily tentative due to the small size of the sample. Nevertheless, since this analysis examined the physician's initial problem formulations in greater detail than has been done in previous research, it may provide some valuable indications as to directions future research in medical problem solving could take.

Analysis of the physician outcome data revealed that what results from the physician's information-processing activity during the early part of the workup is not a unidimensional list of problem formulations, but a structured set of problem formulations which may be described in terms of four features: (1) hierarchical organization, (2) competing formulations, (3) multiple subspaces, and (4) functional

relationships. Of these features, only the second was found to be always present. Thus, it would appear that irrespective of the properties of the case he encounters, the experienced physician always seeks competing explanations for the data obtained during the early part of the workup. Which and how many of the other three features are present, on the other hand, is probably a function: (a) of the properties of the case, and/or (b) of the characteristics of the physician. One goal of future research should be to determine the way in which the structure of a medical case affects the structure of the physician's initial set of initial problem formulations, or, in Newell and Simon's (1972) terms, what properties of the task environment determine the structure of the problem space. To answer this question it will not be sufficient to design cases, such as those used in this study and in other investigations, which are simply a representative sample of the cases encountered by the physician in a given domain. Rather it will be necessary to carefully construct cases along a series of structural dimensions, holding some dimensions constant while varying others. A second, and complementary, goal of future research should be to determine what type of individual difference variables, if any, affect the physician's problem formulation outcomes. There is a wide range of variables that could potentially be

investigated: e.g., amount and type of clinical experience, area of specialization, cognitive style variables, personality traits.

With respect to the processes involved in generating a set of initial problem formulations, the findings of this study suggest: (a) that the major mechanism involved is associative retrieval, and (b) that the mode of representation is primarily verbal. Although the early generation of problem formulations is itself a strategy for dealing with the diagnostic task as a whole, the use of strategies to generate these formulations was not found to be a major characteristic of the physician's information-processing activity during the early part of the workup. Only two strategies were consistently employed by at least half of the physician sample and on at least half of the cases: (1) focusing on diseases of high incidence for the patient's demographic group, and (2) attempting to think of competitors to each formulation generated. Two implications for future research may be drawn from these findings. If the generation of initial problem formulations is largely a process of direct associative retrieval, rather than strategy-guided search, then a major focus of future research should be the investigation of: (a) the organization, or structure, of the physician's store of diagnostic categories in longterm memory, and (b) the properties of the retrieval system

(i.e., the indexing of diagnostic categories in terms of cues) which permits access to this store. On the other hand, it is possible that strategies do play a major role in the generation of problem formulations, but that reliance on introspective data, as was the case in this and other recent studies, is not the means for identifying them. Thus, a second goal of future research should be to attempt to devise tasks which require the subject to externalize the steps in his thinking, and which, because of the properties built into the task by the experimenter, permit inferences about the use of strategies from observations of behavior, i.e., a Bruner, Goodnow and Austin (1956), rather than a Newell, Shaw and Simon (1958), approach to the study of cognitive processes. It will, however, be considerably more difficult to devise an appropriate task for the study of medical problem solving than it was for the Bruner, et al. investigation of concept attainment.

The outcomes of these lines of research will have important implications for the training of medical students. If the generation of problem formulations <u>is</u> found to be primarily a process of associative retrieval, this would imply that a major determinant of the student's acquisition of this skill is the way in which his store of medical knowledge is structured. Thus, future instructional research would need to focus not only on devising specific training in the generation of problem formulations (as was done in

this study), but also on determining how the medical school basic science curriculum should be organized and taught so as to impart to the student a store of medical knowledge that is structured similarly to that of the experienced physician.

### Training of Medical Students in the Generation of Initial Problem Formulations

#### Conclusions

The results of the training experiment support two major conclusions:

1. That a training model consisting of--(1) problemsolving exercises in which films are used to simulate the conditions of the early part of the clinical workup; (2) feedback based on data from a sample of experienced physicians--is an effective means of improving the second-year medical student's skill in generating a set of initial problem formulations.

2. That the training model is just as effective, if not more effective, for second-year students when it provides outcome feedback only, rather than outcome and process feedback.

The analysis of student posttest performance in terms of four variables (CUE, PF, CUE-PF, and R-PF scores) suggests the following conclusions regarding the <u>nature</u> of the training effect: 1. The ability to detect, encode and make at least limited use of the cues presented during the early part of the workup is already well established in the second-year medical student, and thus shows no improvement as a result of the training.

2. The major effect of the training, therefore, lies in the improvement of the student's ability to make use of cues, once obtained, in order to generate a thorough and appropriate set of initial problem formulations.

3. The ability of the second-year medical student to classify cues obtained with respect to problem formulations generated (in a flexible and appropriate manner) does not in itself improve as a result of training. But, his performance on this variable does increase because of the improvement in the scope and appropriateness of his problem formulations.

4. There is no change in the student's ability to identify possible functional relationships among the problem formulations he generates.

#### Implications for Future Research

Given the effectiveness of the "simulation exercises, plus feedback" model as a means of training medical students in the generation of initial problem formulations, one line of future research would be to consider ways of applying the model with respect to other problem-solving skills involved

in the remainder of a clinical workup: namely; the testing of initial problem formulations by further data collection. the revision of initial formulations and the generation of new formulations in light of additional data obtained as the workup progresses, and, ultimately, the making of diagnostic decisions at the close of the workup. Since an essential feature of the physician's activity, after the earliest part of the workup, is the selection of clinical procedures that will provide data to test his problem formulations, it would be necessary to use some other medium than films to simulate the conditions of the remainder of the workup. Booklets with "rub out" answer sheets, of the type employed in the Patient Management Problems developed by McGuire (McGuire and Solomon, 1971), provide an effective (yet relatively low cost) means of simulating sequential decision-making regarding the selection of clinical procedures. Modification of the McGuire format to provide feedback, based on physician performance data, at various key decision points in the workup (e.g., between history and physical, between physical and lab) would be one means to test the effectiveness of the training model with respect to skills other than the generation of initial problem formulations. For example, if one were interested in training which focused on the testing of problem formulations, the training exercise could begin by providing the student with a set of initial formulations to be tested. After he has

completed a data selection sequence (e.g., the history) and recorded his interpretation of the data with respect to the formulations he was given, he could be given feedback: (a) on the data that experienced physicians selected, and (b) the way in which experienced physicians interpreted the data. If, on the other hand, one were interested in providing the student with training in both the generation and testing of problem formulations, it would be possible to construct a more comprehensive simulation exercise, using films with feedback for the early part of the workup, and McGuire type materials with feedback for the remainder of the workup.

A second line of research would be to determine if there are not less expensive means of simulating the early part of the workup than motion picture films of the type used in this experiment. It seems probable, in retrospect, that a set of color slides of the patient, plus a tape recording of the doctor-patient dialogue would be just as effective as a motion picture film, and much less costly. A slide-tape combination would maintain the realism of the audio aspect of the simulated encounter with the patient, but would, of course, reduce the realism of the visual aspect. The question, therefore, is how much does a moving picture of the patient, as compared to a series of still images, contribute to the development of the student's skill in generating problem formulations. Certain types

of visual cues--such as the patient's gait as he enters, the way he sits down, his changes of positions in the chair, the movements of his head and arms, his changes of expression--could not be adequately conveyed by slides. However, examination of both the physicians' and the students' responses reveals that no cues of this type were used to generate problem formulations. The visual cues that the physicians and students did use to generate problem formulations were either essentially static in nature (e.g., the patient's physical build, or his dress) and thus could be easily conveyed by means of slides, or were movements and gestures whose cue properties could be fairly effectively captured in slides (e.g., the patient points to the location of his pain, clutches his abdomen or splints his chest wall with his arm; the patient rests his head on his hand, squints his eyes or exhibits an expression of pain). Thus, for the particular cases used in this study, as well as for a wide range of other possible cases, a moving picture may not play an essential role in conveying visual cues to the learner. Slides would simplify the learner's task by providing still images of the visual cues of relevance to the generation of problem formulations, whereas a film requires the student to detect such cues from the on-going flow of images of the patient. However, the results of the experiment indicate that, at least in the case of second-year medical students, the ability to detect relevant cues on the basis of

naturalistic observation of the patient in motion was <u>already</u> well established; thus, the use of motion picture films was not needed to develop the students' skill in this domain. For first-year medical students, however, a motion picture film might provide needed practice in cue detection, and, therefore, substantially contribute to the training's effectiveness.

A third line of future research that may be proposed pertains to the feedback component of the training model. First, it would be of interest to determine the degree to which feedback contributes to the effectiveness of the model by comparing students' performance under two experimental conditions: (1) simulation exercises with feedback; (2) simulation exercises without feedback. A second question of interest is whether there may be an ordinal interaction between the type of feedback provided ("outcome" versus "outcome and process") and the level of medical knowledge and skill of the student. The results of this experiment indicate that, for second-year medical students, provision of process feedback, in addition to outcome feedback, clearly had no positive effect on the development of problem formulation skills, and may even have had a negative effect, as compared to outcome feedback only. In Chapter V the following explanation was advanced to explain these results. First, it was argued that, having been given outcome feedback, the student had no difficulty in reconstructing what

the physicians' reasoning processes must have been, and thus was able to provide himself with self-generated process feedback. Second, it was proposed that, by providing the student with process feedback which he could generate for himself, the feedback became overly redundant; thus, the student's level of cognitive involvement in the task, as well as his motivation to carry out the task, diminished. It is possible, however, that process feedback would be effective with students at an earlier point in the medical school curriculum. Unlike the second-year student, the first-year student may not have acquired a sufficient level of medical knowledge and skill to be able to reconstruct for himself the processes by which the experienced physician arrives at a given set of problem formulation outcomes. Thus. process feedback materials could be effective in enabling him to understand and assimilate the outcome feedback materials. Replication of the experiment with student entry level as a factor in the design (e.g., third term first-year students and third term second-year students) would be useful to determine whether different types of feedback are appropriate for students having different levels of medical knowledge and skill. The expense of producing one of the training exercises increases considerable by the addition of process feedback: (a) because of the sizable technical cost involved in adding the "think aloud" segments to the standard training film, and (b) because the

collection of the process data from a sample of physicians requires lengthy individual sessions, whereas outcome data only can be easily obtained in group sessions. Thus, unless it can be demonstrated that the addition of process feedback makes the training exercises considerably more effective for some medical student populations, further development of such materials would not be warranted.

Whatever direction is taken by future research on problem-solving instruction (in medicine, or in other domains), two general methodological recommendations may be offered. It is believed that the positive outcome of this experiment (at least with respect to the treatment versus control hypothesis) was due in part to the relatively sizable period of training which the students received. Α series of one to two-hour training sessions over a period of several weeks (rather than the all too frequent single session, 60-minute treatment) would seem to be required in order for a training model (in the area of problem solving) to be properly tested. Secondly, it is believed that the use of multiple dependent measures, each focused on one component of the overall complex of skills that could potentially be affected by the training, (as was the case in this study) is likely to be necessary for training experiments to yield meaningful results. Since some of the skills that are involved in a problem-solving task may improve as a result of training, while others do not, a

battery of quite specific dependent measures would seem to be required in order to determine not only whether the training had an effect but also the precise nature of its effect.

#### Instructional Applications

In conclusion, we would like to suggest that, even without further research and development, the materials produced and tested in this study may have a number of useful applications within the current medical school curriculum.

1. Self-instruction. Each of the simulation exercises could be easily packaged as a self-contained unit (i.e., film cassette, plus instructional booklet, plus response booklet), and the set of such units made available to students for use on an individual basis. The students' responses to the questionnaire indicated that if a "library" of such units were available, most students would make use of it.

2. Group instruction. The materials could also be used in group settings, such as a "focal problems" class. In such a setting it would probably be quite effective to have a group discussion (in which the students compare and criticize the various outcomes they arrived at) prior to presentation of feedback on the physician outcomes. In responding to the questionnaire, many students indicated a preference for using the materials in a group discussion setting, rather than on an individual self-instructional basis.

3. Evaluation. The films produced in this study could be used in designing more effective evaluation instruments for clinically-oriented coursework or clerkships. Although the films provide a simulation of only the early part of the workup, they could be combined with booklets of additional clinical findings (i.e., further history, plus physical and lab data) in order to evaluate a wide range of clinical competencies.

The results of this study indicate that either version of the training materials could be an effective instructional tool for improving second-year medical students' skill in generating initial problem formulations. Although neither version of the materials was found to be conclusively more effective than the other, there was some evidence to suggest that, at least for second-year students, the outcome feedback version (with certain modifications recommended in Appendix J) is likely to be most appropriate for instructional use. It is hoped, moreover, that medical educators will find ways of adapting, expanding and improving upon the materials that were developed in this study.

. .

·

# REFERENCES

#### REFERENCES

- Barrows, H. S. & Bennett, K. The diagnostic (problem solving) skill of the neurologist. <u>Archives of</u> <u>Neurology</u>, 1972, 62, 273-277.
  - Bartlett, F. C. Thinking. New York: Basic Books, 1958.
  - Broadbent, D. E. <u>Perception and communication</u>. New York: Pergamon Press, 1958.
  - Bruner, J. S. The process of education. New York: Vintage Books, 1960.
  - Bruner, J. S. <u>Toward a theory of instruction</u>. Cambridge, Mass.: Belknap Press, 1966.
  - Bruner, J. S., Goodnow, J. J., & Austin, G. A. <u>A study</u> of thinking. New York: Wiley, 1956.
  - Campbell, D. T. & Stanley, J. C. <u>Experimental and Quasi-</u> <u>experimental designs for research</u>. Chicago: Rand McNally, 1969.
  - Chamberlin, T. C. The method of multiple working hypotheses (1890). Reprinted in Science, 1965, <u>148</u>, 754-759.
  - Collins, A. M. & Quillian, M. R. Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 240-247.
  - Cornfield, J. & Tukey, J. W. Average values of mean squares in factorials. <u>Annals of Mathematical Statistics</u>, 1956, <u>27</u>, 907-949.
  - Craig, R. M. Recent research on discovery. <u>Educational</u> Leadership, 1969, 26, 501-508.
  - Cronbach, L. J. Coefficient alpha and the internal structure of tests. <u>Psychometrika</u>, 1951, <u>16</u>, 297-334.
  - Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. Theory of generalizability: A liberalization of reliability theory. <u>British Journal of Statistical Psychology</u>, 1963, 16, 137-163.

- Dewey, J. Experience and education. New York: Collier Books, 1963. (Originally published 1938.)
- Dewey, J. Logic: The theory of inquiry. New York: Holt, 1938.
- Ebel, R. L. Estimation of the reliability of ratings. <u>Psychometrika</u>, 1951, <u>16</u>, 407-424.
- Egan, D. E. & Greeno, J. G. Acquiring cognitive structure by discovery and rule learning. <u>Journal of Edu-</u> cational Psychology, 1973, 64, 85-97.
- Elstein, A. S., Kagan, N., Shulman, L. S., Jason, H., & Loupe, M. J. Methods and theory in the study of medical inquiry. <u>Journal of Medical Education</u>, 1972, <u>47</u>, 85-92.
- Engel, G. L. The deficiency of the case presentation method of clinical teaching. <u>New England Journal of Medicine</u>, 1971, 284, 20-24.
- Finn, J. D. Univariate and multivariate analysis of variance and covariance: A Fortran IV program. Office of Research Consultation, College of Education, Michigan State University, Occasional Paper No. 9, 1970.
- Gagné, R. M. <u>The conditions of learning</u>. (2nd edition). New York: Holt, Rinehart and Winston, 1970.
- Gagné, R. M. Instruction based on research in learning. Engineering Education, 1971, 61, 519-523.
- Glass, G. V., Peckham, P. D., & Sanders, J. R. Consequences of failure to meet assumptions underlying the fixed effects analysis of variance and covariance. <u>Review of Educational Research</u>, 1972, <u>42</u>, 237-288.
- Glendening, L. POSTHOC: A Fortran IV program for generating confidence intervals using either Tukey or Scheffé multiple comparison procedures. Office of Research Consultation, College of Education, Michigan State University, Occasional Paper No. 20, 1973.
- Gordon, M. J. Heuristic training for diagnostic problem solving among advanced medical students. Unpublished Ph.D. dissertation, Michigan State University, 1973.

- Hammond, K. R. & Kern, F. J. <u>Teaching comprehensive</u> <u>medical care</u>. Cambridge, Mass.: Harvard University Press, 1959.
- Hammond, K. R. & Summers, D. A. Cognitive control. Psychological Review, 1972, 79, 58-67.
- Hammond, K. R., Summers, D. A., & Deane, D. H. Negative effects of outcome-feedback in multiple-cue probability learning. Institute of Behavioral Science, University of Colorado, Program of Research on Human Judgment and Social Interaction, Report No. 148, 1972.
- Harvey, A. M. & Bordley, J. <u>Differential diagnosis</u>. (2nd edition). Philadelphia: Saunders, 1970.
- Harvey, A. M., Johns, R. J., Owens, A. H., & Ross, R. S., (Eds.). The principles and practice of medicine. (18th edition). New York: Appleton-Century-Crofts, 1972.
- Holsti, O. R. <u>Content analysis for the social sciences and</u> <u>humanities</u>. Reading, Mass.: Addition-Wesley, 1969.
- Hoyt, C. J. Test reliability estimated by analysis of variance. Psychometrika, 1941, 6, 153-160.
- Kessel, F. S. The philosophy of science as proclaimed and science as practiced: "Identity" or "dualism"? American Psychologist, 1969, 24, 999-1005.
- Kleinmuntz, B. The processing of clinical information by man and machine. In B. Kleinmuntz (Ed.), Formal representation of human judgment. New York: Wiley, 1968.
- Lewy, A. & McGuire, C. A study of alternative approaches in estimating the reliability of unconventional tests. Paper presented at the Annual Meeting of the American Educational Research Association, 1966.
- Luchins, A. S. & Luchins, E. H. New experimental attempts at preventing mechanization in problem solving. Journal of Genetic Psychology, 1950, <u>42</u>, 279-297.
- Lusted, L. B. & Stahl, W. R. Conceptual models of diagnosis. In J. A. Jacquez (Ed.), <u>The diagnostic process</u>. Proceedings of a conference sponsored by the Biomedical Data Processing Training Program, University of Michigan, 1963.

- Mandler, G. Organization and memory. In K. W. Spence and J. T. Spence (Eds.), <u>The psychology of learning</u> and motivation. New York: Academic Press, 1967.
- Marshall, R. L. Self-generated complexity, cognitive abilities and strategies as determiners of clue utilization in problem-solving. Unpublished Ph.D. dissertation, University of California at Berkeley, 1971.
- McGuire, C. H. & Solomon, L. M. <u>Clinical simulations:</u> <u>Selected problems in patient management and illus-</u> <u>trations</u>. New York: Appleton-Century-Crofts, 1971.
- Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-97.
- Miller, G. A., Galanter, E., & Pribram, K. H. <u>Plans and</u> <u>the structure of behavior</u>. New York: Holt, Rinehart and Winston, 1960.
- Miller, G. E. (Ed.). <u>Teaching and learning in medical</u> <u>schools</u>. Cambridge, Mass.: Harvard University Press, 1962.
- Neisser, U. <u>Cognitive psychology</u>. New York: Appleton-Century-Crofts, 1967.
- Newell, A., Shaw, J. C., & Simon, H. A. Elements of a theory of human problem solving. Psychological Review, 1958, 65, 151-166.
- Newell, A. & Simon, H. S. <u>Human problem solving</u>. Englewood Cliffs, N.J.: Prentice-Hall, 1972.
- Reader, G. G. & Goss, M. E. <u>Comprehensive medical care</u> and teaching. Ithaca, N.Y.: Cornell University Press, 1967.
- Rimoldi, H. J. A. Evaluation and training of clinical diagnostic skills. Psychometric Laboratory, Loyola University, Publication No. 41, 1963.
- Scheffé, H. <u>The analysis of variance</u>. New York: Wiley, 1959.
- Schwartz, S. H. & Simon, R. I. Information processing and decision making in medical diagnosis. Paper presented at a symposium on Health Sciences and the Systems Approach, Wayne State University, 1970.

- Shulman, L. S. Psychology and mathematics education. In E. G. Begle (Ed.), <u>Mathematics Education</u>. Chicago: University of Chicago Press, 1970.
- Shulman, L. S. & Keisler, E. R. (Eds.). <u>Learning by</u> <u>discovery: A critical appraisal</u>. Chicago: Rand <u>McNally</u>, 1966.
- Shulman, L. S., Loupe, M. J., & Piper, R. M. <u>Studies of</u> <u>the inquiry process</u>. Educational Publication Services, College of Education, Michigan State University, 1968.
- Sox, H. C., Sox, C. H., & Tompkins, R. K. The training of physician's assistants. <u>New England Journal of</u> Medicine, 1973, 288, 818-824.
- Sprafka, S. A. The effect of hypothesis generation and verbalization on certain aspects of medical problem solving. Unpublished Ph.D. dissertation, Michigan State University, 1973.
- Tulving, E. & Donaldson, W. (Eds.). Organization of memory. New York: Academic Press, 1972.
- Twelker, P. A. Simulation and media. In P. J. Tansey
  (Ed.), Educational aspects of simulation. New
  York: McGraw Hill, 1971.
- Wason, P. C. On the failure to eliminate hypothesis: A second look. In P. C. Wason & P. N. Johnson-Laird (Eds.), <u>Thinking and reasoning</u>. Baltimore: Penguin Books, 1968.
- Ways, P. O., Baker, T., Finhilstein, P., Fiel, N. J., & Jones, J. W. <u>The problem-oriented record: A self-instructional unit</u>. College of Human Medicine, <u>Michigan State University</u>, 1972.
- Ways, P. O., Loftus, G., & Jones, J. M. Focal problem teaching in medical education. Journal of Medical Education, 1973, 48, 565-571.
- Weed, L. L. <u>Medical records, medical education, and patient</u> <u>care</u>. Cleveland, Ohio: Press of Case Western Reserve University, 1969.
- Winer, B. J. <u>Statistical principles in experimental design</u>. New York: McGraw Hill, 1962.

- Wintrobe, M. M., et al. (Eds.). <u>Harrison's Principles of</u> <u>Internal Medicine</u>. (6th edition). New York: <u>McGraw Hill</u>, 1970.
- Wortman, P. M. Medical diagnosis: An information processing approach. <u>Computers and Biomedical Research</u>, 1972, <u>5</u>, 315-328.
- Wortman, P. M. & Kleinmuntz, B. The role of memory in information-processing models of problem solving. Unpublished paper, Department of Psychology, Duke University, undated.

# APPENDICES

# APPENDIX A

.

.
#### CASE OUTLINE FOR FILM 1:

#### "A 21-year-old College Senior"

A. Written sheet:

sex: Male
age: 21
occupation: college student (senior)
temperature: 99°

B. Verbal dialogue:

#### Complaints

#### Attributes

- 1. weakness and exhaustion
- --feeling increasingly weak and exhausted for sometime (about 2 months)
- --so bad during last couple weeks that he can hardly walk across campus
- --lives in 3rd floor apartment, and by the time he climbs to the top of the stairs he is completely exhausted and has to flop down on the couch
- --after exertion his legs feel wobbly and weak
- -- no shortness of breath
- --began about 4-5 months ago
- --below beltline, on right side only (show)
- --heavy, sharp pains, like a big heavy rock
- --pain comes on in the evening, about 8-9 o'clock and usually lasts until about midnight when he falls asleep; when he wakes up the pain is gone

2. abdominal pain

		he hasn't noticed anything (e.g., certain foods or activities) that make the pain better or worse
		he sometimes takes aspirin but this does not seem to reduce the pain
		in the beginning, the pain was irregular (it would come for a few nights and then go away for a week); but during the last couple months it has gotten worse (more frequent and more intense)
3.	weight loss	he's lost about 25 pounds over the past 3-4 months
		general loss of weight
		he's been eating normally, in fact, he's been eating more than usual (regular meals, plus snacks between meals)
4.	diarrhea	began about 3 months ago (about a month after he first noticed the pain)
	•	at first, 2 stools a day, more recently 3-4 stools a day (in the past, he normally had 1 stool a day)
		<pre>stools are soft, loose, but    not watery</pre>
		has noticed quite a lot of mucous in stools, and some- times some blood, and pieces of food

## Other

- 5. No vomiting.
- 6. Nothing special has happened during the past few months (no school or social problems). He's always worked very

hard, plans his study habits and is very organized about his work. He gets good grades, but has to work hard for them. The exhaustion and pain are getting in his way, making it difficult for him to work and he's worried about being able to finish the term and graduate in June.

C. Nonverbal cues:

thin, appears tired intelligent, well-dressed answers questions carefully and conscientiously appears tense, but not really nervous

NOTE: The patient intially present the complaint of exhaustion. When asked if he has any other problems, he mentions the pain. Further questions by the doctor elicit items 3-6. APPENDIX B

#### PROCESS CHECKLIST

Subject \_\_\_\_\_ Film \_\_\_\_\_

Check as many items as apply. Do not check items that you consider to be praise-worthy, or which describe your clinical approach in general. Check only those items which characterize your thinking while viewing this film.

- 1. As the patient described his complaints, this information elicited a sort of "mental list" of possible problem formulations.
- 2. In attempting to come up with problem formulations, I made an effort to think of the most serious (lifethreatening) types of diseases that the patient might have.
- 3. On the basis of the information presented in the film, I made some quick "rule-outs" of several problem formulations.
- 4. In attempting to come up with problem formulations, I focused on pathophysiological processes (i.e., what aspect of physiology is disturbed, and what could cause this disturbance?)
- 5. I waited until I had some data on each of the patient's major complaints before attempting to look for interrelationships among the symptoms.
- 6. I tried early on to form a general impression of the patient (his personality, intelligence, background, etc.) that would help me to judge whether he was giving me accurate, objective history information.
- 7. In attempting to arrive at problem formulations, one or more sorts of "mental images" came to mind.
- 8. One particularly salient piece of data immediately brought to mind one or more problem formulations.
- 9. Given the patient's major complaints, I tried to think of as many different organic causes as possible.
- 10. I assumed that the patient's problem was organic, unless confronted with evidence to the contrary.

- \_\_\_\_11. One of the first things I tried to do was to localize the patient's problem in terms of some organ system.
- 12. In attempting to come up with problem formulations, I tried to think of those illnesses that have a high statistical incidence for persons of the patient's sex, age group, occupational group, etc.
- \_\_\_\_13. In attempting to come up with problem formulations, I gave a great deal of weight to the first complaint mentioned by the patient.
- 14. In attempting to arrive at problem formulations, I focused on a couple pieces of data that appeared to be most critical, and paid less attention to the other pieces of data.
- \_\_\_\_15. It was the combination of the patient's major complaints that led me to think of one or more problem formulations.
- \_\_\_\_16. One of the first things I tried to do was to determine whether the patient's problem was organic or psychogenic.
- \_\_\_\_17. I tried to come up with at least one problem formulation that would account for <u>all</u> of the data presented.
- \_\_\_\_18. One of the first things I tried to do was to determine whether the patient's problem was acute or chronic.
- 19. In attempting to come up with problem formulations, I focused primarily on those illnesses that are most common, given the patient's chief complaints and demographic characteristics.
- 20. In thinking of problem formulations, I paid a great deal of attention to nonverbal characteristics of the patient, e.g. his build, posture, facial expression, gestures, emotional state, etc.
- 21. In attempting to arrive at problem formulations, it helped me to try to visualize (i.e., to form some sort of "mental image" of) the anatomical location of the problem.
- \_\_\_\_22. As each piece of data was presented I tried to think of how it might be related to the other pieces of data.

- \_\_\_\_23. As I observed the patient and listened to his complaints, I recalled another patient (or patients) that I had seen before.
- \_\_\_\_24. As soon as a problem formulation came to mind, I made an effort to think of other problem formulations that need to be considered.
- 25. In attempting to come up with problem formulations, I paid attention primarily to what the patient was saying.

# APPENDIX C

.

### TRAINING MATERIALS

This Appendix contains the following materials:

 The "Introduction" section of the Instructional Booklet (Treatment I).

2. The "Film 1" section of the Instructional Booklet (Treatment I).

3. The "Self-Evaluation Checklist" for Film 1 (both treatment groups).

# INTRODUCTION

•

This instructional package focuses on one aspect of the physician's activity in conducting a clinical workup: namely, the generation of a set of <u>initial problem formulations</u> during the first minutes of an encounter with a patient. The materials have been designed to provide you with the opportunity to practice generating initial problem formulations for a variety of medical cases. For each case, an instructional sequence consisting of three basic components will be followed: (1) you will view a film of the first 4-6 minutes in a doctor-patient encounter; (2) having viewed the film, you will record the problem formulations you have generated and write a tentative assessment; (3) you will be provided with "feedback materials" which describe the problem formulations and tentative assessments generated by a group of experienced physicians who have viewed each film.

233

-1-

WHAT IS "AN INITIAL PROBLEM FORMULATION"?

During the first minutes of a clinical workup, as the patient's presenting complaint(s) are elicited, the experienced physician usually generates a set of <u>initial problem formulations</u>. These initial formulations serve two functions:

- They are a set of mental categories under which the physician organizes and classifies the data obtained during the first minutes of the workup.
- (2) They are a set of tentative hypotheses which he will subsequently test by collecting further data during the remainder of the workup.

The <u>specificity</u> of an initial problem formulation will depend on the adequacy of the data obtained during the first minutes of a workup.

- --An initial problem formulation may be highly general, e.g., "organic disorder," "psychological problem,"
- --It may refer to an organ system, e.g., "renal disorder,"
- --It may refer to a disease mechanism, e.g., "infection,"
- --It may refer to both an organ system and a disease mechanism, e.g., "renal infection,"
- --It may be a highly specific diagnostic label, e.g., "glomerulonephritis."

The <u>number</u> of initial problem formulations a physician generates will depend on the number of distinct complaints the patient presents and the degree to which these complaints are consistent with a single or multiple formulations.

In actual practice, the physician does not record his initial problem formulations. Rather he mentally stores them, evaluates them with respect to the data he collects, reformulates them or generates new formulations as needed, and, at the end of the workup, records those problem formulation he has retained as entries on a Problem-Oriented Record. In using this instructional package, however, your task will end with the recording of the initial problem formulations you have generated after viewing a 4-6 minute film of an encounter with a patient. COMPONENTS OF AN INITIAL PROBLEM FORMULATION.

Each of your initial problem formulations should include two components, and, in some cases, may include a third component:

- 1. a problem formulation title,
- 2. a list of cues,
- 3. (optional) a list indicating more specific diagnostic possibilities under consideration, and the cues of particular relevance to each.
- A <u>problem formulation title</u> is a label, having potential diagnostic and/or management implications, under which you are able to group elements data (or cues) presented in the filmed interview. It should be stated at a level of specificity that is appropriate to the available data.
- A <u>cue list</u> should include all elements of data that are relevant to the problem formulation title under which they are listed. The list may include both information reported verbally by the patient and nonverbal cues which you observe.

For some problem formulations, there may be particular cues which suggest

one or several <u>more specific diagnostic possibilities</u>. For example, among the cues which have led you to generate the problem formulation of "renal disorder" there may be several cues which point to "glomerulonephritis" as one diagnostic possibility to be considered. In this case, you would record "glomerulonephritis" (plus the cue(s) of *particular* relevance to it) as a diagnostic possibility under the problem formulation "renal disorder."

#### WRITING A TENTATIVE ASSESSMENT

After you have recorded your initial problem formulations for a case, you will be asked to write a brief paragraph giving your <u>tentative</u> <u>assessment</u> of these formulations. Your tentative assessment for each case will discuss the *set* of initial problem formulations you have generated after a 4-6 minute encounter with a patient. It should indicate:

-5-

- --how well substantiated you consider each of your problem formulations to be on the basis of the data obtained thus far;
  --whether you anticipate that the patient has a single illness that will account for his various problems, or that he has multiple disorders;
- --whether you consider there to be any relationships among your problem formulations. For example, you may consider one problem formulation to be secondary to, superimposed on, or contributing to, etc. some other formulation.

You will note that in using this instructional package the format to be followed, in recording problem formulations and tentative assessments, differs somewhat from the usual Problem-Oriented Record format.

In the usual Problem-Oriented Record:

an assessment is written for each problem formulation, and discussion of various diagnostic possibilities is included in each assessment.

In this instructional package:

you are asked to list specific diagnostic possibilities (if any) under each problem formulation, and to write a single tentative assessment summarizing all of the problem formulations you have generated for a case.

## <u>A</u> <u>reminder</u>....

As you view the films and attempt to generate a set of initial problem formulations, keep in mind that these initial formulations are *tentative* hypotheses which you would want to investigate more thoroughly if you were to continue the workup beyond the first 4-6 minutes presented in the film.

-6-

### THE INSTRUCTIONAL MATERIALS

#### The Films

Color films will be used to simulate your encounter with eight patients. Each of these films presents a "physician's eye view" of the first 4-6 minutes in an office visit with a new patient. Throughout the interview the film focuses on the patient; the physician's voice is heard but he is never seen. While viewing the film, you should attempt to put yourself in the role of the physician. Pretend that the patient is sitting in front of you and talking to you.

-7-

### The Response Booklet

The Response Booklet is divided into eight sections. There is one section for each of the eight films you will view. Each section contains the following materials: (1) a set of response sheets on which you will record the problem formulations you have generated; (2) a sheet on which you will write your tentative assessment; and (3) a selfevaluation checklist to be filled out at the end of the instructional sequence.

### The Feedback Materials

The feedback materials summarize the initial problem formulations and tentative assessments generated by a group of eight experienced physicians who have viewed the films. The purpose of these materials is to provide you with a means of comparing your own performance on each case to that of experienced physicians.

THE INSTRUCTIONAL SEQUENCE

For each of the filmed interviews, the same instructional sequence will be followed. The steps in the sequence are summarized below. This summary is intended to provide you with an overview of the instructional sequence. Complete instructions for each step will be repeated, as appropriate, throughout the INSTRUCTIONAL BOOKLET.

STEP 1. You will read the "nurse's sheet" for the patient in the film.

This sheet indicates the patient's name, sex, age, occupation, and temperature (taken orally by the nurse).

STEP 2. You will view the film of the 4-6 minute interview of the patient.

As you view the film, you should generate a set of initial problem formulations.

STEP 3. You will record the problem formulations you have generated, and write a brief tentative assessment.

STEP 4. You will be provided with feedback materials describing the performance of the group of experienced physicians.

a. You will be provided with "Feedback Sheet 1."

This sheet presents the major problem formulations generated by the physicians; i.e., those formulations generated by all or nearly all physicians who viewed the film. b. You will view the film a second time.

This viewing of the film will provide you with the opportunity to repeat your encounter with the patient. As you view the film, attempt to reconstruct in your own mind the reasoning process which led the physicians to generate the problem formulation(s) listed on Feedback Sheet 1.

c. You will be provided with "Feedback Sheet 2."

This sheet has two sections.

The first section presents <u>additional</u> problem formulations generated by some of the physicians who viewed the film.

The second section (entitled "Summary") describes the physicians' tentative assessments.

STEP 5. You will fill out a self-evaluation checklist designed to aid you in comparing your performance to that of the experienced physicians. GUIDELINES FOR COMPLETION OF THE PROBLEM FORMULATION RESPONSE SHEETS

- 10 -

In reading these guidelines, you should refer to the two sample problem formulation sheets on pages 13-16. Both of these sheets apply to the same patient.

1. At the top of each response sheet, list the title of one problem formulation you have generated. Underneath, in the space provided, list the cues (i.e., all pieces of relevant data) for this formulation.

2. In listing the cues, try to record, as closely as possible, the words used by the patient, or, in the case of nonverbal cues, your actual observation. If you make any interpretations or inferences based on a cue, put these in parentheses. For example: "swollen fingers (edema?)"

3. List both "positive" cues (i.e., cues that tend to confirm a problem formulation) and "negative" cues (i.e., cues that tend to disconfirm a problem formulation). If you consider a cue to be "negative" for a problem formulation, indicate this by writing "(neg.)" in front of the cue. (See the examples on page 15.)

4. A cue may be listed under more than one problem formulation.
A cue which is listed as "positive" for one problem formulation may be listed as "negative" for some other problem formulation.

5. For some of your problem formulations, there may be certain cues which suggest one or several more specific diagnostic possibilities. If this occurs, then on the *back* of the problem formulation response sheet, you are to list:

- (a) the title of each diagnostic possibility under consideration for that problem formulation;
- (b) next to each possibility, in the space provided,

the cues that are of *particular* relevance to it. (See the example on pages 13-14.)

6. Write legibly and avoid abbreviations.

•

7. If you want to take notes while viewing a film, you may do so. Use a sheet in the Response Booklet for note-taking, and write "NOTES" at the top of the sheet.

The following pages present response sheets for a sample patient, including (a) two examples of initial problem formulations, and (b) an example of a tentative assessment based on these formulation. Take several minutes to look over these sheets and to review the preceding "Guidelines."

You can also look back over any of the materials presented up to this point in this booklet.

243 / 1

Film<u>Example</u> Problem formulation title:<u>renal disorder</u>

245

CUE LIST
swollen fingers + ankles (edema?), I week
loss of appetite + nausea, 3.4 days
vomiting, 1Yz days
darker urine (hematuria?), 2days
weakness + fatigae, (week
headache & (hypertension?)
dizziness
age 22
J
·
More specific diagnostic possibilities

(if any) are to be listed on the back of the sheet.

÷

More specific diagnostic possibilities which you have considered for this problem formulation (if any)

Title	Cues of particular relevance
1. glomeralo nephritis	<u>swollen fingers 4 antles (edema?</u> <u>darker urine (hematuria?)</u> <u>sore throat 2 weeks ago</u> -
2.	
3.	
4.	
5.	
<u></u> ń.	

Film Example.

Problem formulation title: <u>Congestive heart failure</u>

CUE LIST swollen antles (edema?), Iweek loss of appetite & nausea, 3.4 days Weakness + fatigue, I week Z (hypertension?) headache dieziness (neg.) no history of cardiac disease (nec.) age 22 More specific enagnostic possibilities

if any! une to be listed on the back

of the sheet.

	Title	Cues of particular relevance
1.		
2.		
3.		
4.		
5.		
6.		

•

More specific diagnostic possibilities which you have considered for this problem formulation (if any)

Film Example

#### TENTATIVE ASSESSMENT

The patient's symptoms point strongly to some type of renal disorder. The fact that she had a sore throat 2 weeks ago (possibly a strep. infection), and the cues indicating probable edema and hematuria, lead me to consider the possibility of glomerulo mephritis. Many of her symptoms would also be consis-tent with congestive heart failure. In a person age 22 with no previous history of cardiac disease, congestive heart failure as a primary problem would be guite unlikely. But, it is possible that she has congestive heart failure secondary Given the recent onset of her symptoms, the fact that she is young and appears generally healthy, I would anticipate that she has a single illness : a renal disorder such 03 giomerula nephritis, either with or without congestive heart failure as a implication.

If nore spice is defauld, continue on back of page.





FILM 1

-21-

STEP 1. Here is the nurse's sheet for the patient in film 1.

Name: Bo	b Renni	inger	
Sex: <u>M</u>		J Age: <b>21</b>	
Occupation:_	college	student	(senior)
Temperature:	99° F		

STEP 2. The film of the 4-6 minute interview with the patient will now be presented.

While viewing the film, you should generate a set of initial problem formulations which you would want to investigate more thoroughly if you were to continue the workup beyond the first 4-6 minutes presented in the film.

(PRESENTATION OF THE FILM)

STEP 3. Turn to the section of the Response Booklet for film 1.

Record the problem formulation you have generated. Fill out one response sheet for each problem formulation.

You may refer back to the "Guidelines" for completion of these sheets, on pages 10-11, if you wish.

(RECORD PROBLEM FORMULATIONS)

After you have recorded your problem formulations, write a brief paragraph giving your <u>tentative</u> <u>assessment</u> of the case. Your assessment should indicate:

- --how well substantiated you consider each of your problem formulations to be on the basis of the data obtained thus far;
- --whether you anticipate that the patient has a single illness that will account for his various problems, or that he has multiple disorders;
- --whether you consider there to be any relationships among your problem formulations. For example, you may consider one problem formulation to be secondary to, superimposed on, or contributing to, etc. some other formulation.

(WRITE TENTATIVE ASSESSMENT)

After you have written your tentative assessment, go on to the next page.

FILM 1

# FILM 1

-23-

STEP 4. You will now be provided with feedback on the performance of the group of experienced physicians who viewed this film.

Turn to the next page and read Feedback Sheet 1.

Check your response sheets to see if they include the major problem formulation(s), listed on Feedback Sheet 1, which were generated by all or nearly all of the physicians.

## Film 1 Feedback Sheet 1

Major Problem Formulations

All physicians who viewed this film generated the following problem formulation:

## GASTROINTESTINAL DISORDER.

Under this formulation, all physicians listed the possibility of inflammatory bowel disease, such as:

ULCERATIVE COLITIS, or REGIONAL ENTERITIS/ILEITIS.

The following table presents the cues listed as relevant to GI DISORDER in general, and to ULCERATIVE COLITIS or REGIONAL ENTERITIS/ILEITIS in particular.

GI DISORDER: ULCERATIVE COLITIS, or REGIONAL ENTERITIS/ILEITIS

GI disorder pain in right lower quadrant of abdomen, for 4 months occurs in evening, lasting several hours not relieved by aspirin or Darvon not related to foods diarrhea: increase in number of stools, from 1 to 4-5/day, over 3 month period mucous in stools blood in stools pieces of food in stools weight loss, 25 lbs. in 1-2 months good appetite, eating more than usual extreme fatigue and weakness, for 2 months no vomiting Ulcerative colitis, or regional enteritis/ileitis diarrhea blood and mucous in stools weight loss of 25 lbs., with good appetite age 21 college senior: under academic stress concerned about keeping up with studies

FILM 1

-25-

STEP 4. Feedback (continued)

The film will now be presented for a second time. As you view the film, attempt to reconstruct in your own mind the reasoning process which led the physicians to generate the problem formulation(s) listed on Feedback Sheet 1.

(PRESENTATION OF THE FILM)

Now turn to the next page and read Feedback Sheet 2.

Compare your problem formulations to those listed on the feedback sheets. Compare your tentative assessment to the physicians' assessments described in the "Summary" on Feedback Sheet 2.

### Film 1 Feedback Sheet 2

## Additional Problem Formulations

Under the formulation of GI DISORDER, most physicians listed, in addition to inflammatory processes (ulcerative colitis, or regional enteritis/ileitis), two other possibilities:

INTESTINAL MALIGNANCY,

## PSYCHOGENIC PROBLEM.

The cues listed as particularly relevant to each of these possibilities are presented below.

GI DISORDER: INTESTINAL MALIGNANCY; PSYCHOGENIC PROBLEM

<u>GI disorder</u> (See cues listed on Feedback Sheet 1.)	
Intestinal malignancy blood in stools weight loss of 25 lbs. extreme fatigue (neg.) age 21	
<u>Psychogenic problem</u> initial impression: patient appeared tense and anxious college senior: under academic stress concerned about keeping up with studies (neg.) blood in stools (neg.) weight loss of 25 lbs. with good appetite (neg.) patient appeared to be frank and objective in reporting symptoms (neg.) reported no major difficulties with school or personal life	

Film 1 Feedback Sheet 2 (Continued)

In addition to the formulation of GI DISORDER, some physicians generated the following problem formulations:

ANEMIA, DUE TO GI BLOOD LOSS

CARDIOVASCULAR PROBLEM,

DIABETES MELLITUS.

ANEMIA, DUE TO GI BLOOD LOSS

extreme fatigue, for 2 months especially on exertion weakness in leg muscles blood in stools onset of diarrhea (with blood?) preceded onset of fatigue by 1-2 months

### CARDIOVASCULAR PROBLEM

extreme fatigue, for 2 months especially on exertion weakness in leg muscles (neg.) no shortness of breath

DIABETES MELLITUS

extreme fatigue and weakness, for 2 months weight loss, 25 lbs. in 1-2 months good appetite, eating more than usual
### Film 1 Feedback Sheet 2 (Continued)

<u>Summary</u> (of the physicians' tentative assessments)

All physicians stated that they anticipate a single illness: most probably some type of gastrointestinal disorder. All indicated that, given the patient's symptoms and his age, inflammatory bowel disease, such as ulcerative colitis or regional enteritis/ileitis, is the most likely type of GI disorder for him to have. An alternative hypothesis, intestinal malignancy, was considered by most physicians, but was judged to be less likely than the inflammatory processes under consideration. Most physicians also gave consideration to the possibility of a GI disorder of psychogenic origin. However, they tended to conclude that the patient probably has an organic GI disorder, such as ulcerative colitis, which may be aggravated by psychological factors (e.g., academic stress), rather than a problem that is primarily psychogenic in nature.

Due to the patient's extreme fatigue and weakness, some physicians generated the formulation of anemia, due to blood loss from the GI tract, and several generated the formulation of cardiovascular problem. They tended to conclude that a cardiovascular problem (unrelated to his GI difficulties) is quite unlikely, whereas anemia (secondary to intestinal inflammation or malignancy) is very possible.

Several physicians indicated that diabetes mellitus, is another possibility to explore, although on the basis of the available data it does not appear to be very likely.



# FILM 1

STEP 5. Now turn to the self-evaluation checklist at the end of the Response Booklet section for film 1.

(FILL OUT SELF-EVALUATION CHECKLIST)

### SELF-EVALUATION CHECKLIST

#### INSTRUCTIONS

The checklist is designed to aid you in evaluating your own problem formulation performance as compared to that of the experienced physicians.

Part A of the checklist presents the titles of all problem formulations (and diagnostic possibilities) generated by the physicians. Part B presents the statements regarding possible relationships between problem formulations that appeared in the physicians' tentative assessments.

Place a check next to each item in the checklist which corresponds to one of your own responses. In order to check an item, there does not have to be an exact correspondance in wording; a general equivalence in meaning is sufficient.

(FILL OUT THE CHECKLIST)

If you have checked	You may consider that the degree of correspondance between your own performance and that of the experienced physicians is
All items marked (*), and some (or all) of the other items;	high
All items marked (*), and none of the other items; OR Some items marked (*), and some of the other items;	moderate
None of the items marked (*), and some (or none) of the other items.	low

Note: It should be borne in mind that the checklist is not necessarily exhaustive. If a larger group of physicians had viewed the film, it is possible that the checklist would have included some additional items. If your own set of problem formulations and/or tentative assessment included items that do not appear in the checklist, you cannot evaluate them by means of the checklist, but this does not necessarily mean that they are inappropriate.

# SELF-EVALUATION CHECKLIST

Before filling out the checklist, read the instructions on the opposite page.

### CHECKLIST FOR FILM 1

<u>Part A</u> :	Problem formulations, and diagnostic possi- bilities
1.	GI disorder*
	<pre>a. ulcerative colitis, and/or regional enteritis*</pre>
	b. intestinal malignancy
	c. psychogenic problem
2.	anemia due to GI blood loss
3.	cardiovascular problem
4.	diabetes mellitus
<u>Part B</u> :	Relationships between problem formulations
1.	anemia secondary to ulcerative colitis/regional enteritis, or intestinal malignancy

Note: An (\*) indicates that the item was included in the responses of all or nearly all of the physicians.

APPENDIX D

.

#### Film 8

Place a check in front of each cue that you remember seeing or hearing in the filmed interview.

has been drinking a lot recently 1. 2. her voice quavers 3. has recently noticed some blurring of vision abdominal pain is localized in right lower quadrant 4. 5. is leaning forward, clutching abdomen 6. 7. 8. 9. has vaginal discharge, along with itching female abdominal pain 9. no fever -10. obese 11. vaginal itch has re-occurred several times during past year 12. feels nausea after eating fatty foods 13. has been "tired & edgy" recently 14. vaginal itch re-occurred 2 days ago 15. boyfriend used contraception 16. has had intermittant headaches recently sudden onset of abdominal pain this a.m. 18. urine has had darker color recently 19. breasts sore and swollen, for 1 week 20. stopped taking birth control pills 1 year ago, due to vaginal itch 21. pain radiates from abdomen to back 22. burning on urination, 1 week 23. has been feeling "funny" recently 24. no recent change in bowel movements 25. has noticed shortness of breath recently 26. abdominal pain is diffuse 27. projectile vomiting this a.m. 28. pain relieved by lying down 29. has been eating a lot of candy and cookies recently 30. has felt "very depressed" recently 31. last menses 8 weeks ago 32. no bowel movement this a.m. 33. menstrual cycle is usually irregular 34. has felt dizzy recently 35. student nurse 36. anxious manner & tone of voice 37. nausea for several days 38. abdominal pain is intermittant and crampy 39. has noticed no recent weight change 40. vomiting began after breakfast this a.m.

- 42. increased appetite recently
- 43. skin rash on hands for several days
- 44. has broken off with boyfriend 1 month ago
- 45. has noticed recent increase in frequency of urination
- 46. age 19
- 47. sexually active
- 48. steady increase of abdominal pain
- 49. increased thirst recently
- 50. abdominal pain relieved by vomiting
- 51. was using contraceptive foam after stopped taking birth control pills
- 52. has been eating irregular meals recently
- 53. has pain on intercourse
- 54. has not been able to concentrate recently
- 55. had cheese sandwich & milk at midnight, 6 hours before onset of abdominal pain & vomiting
- 56. rests her head on her hand
- 57. vomits thin yellow liquid
- 58. complains of frequent belching
- 59. her condition is stable at present
- 60. admits anxiety about pregnancy
- 61. sudden chill last night
- 62. usually has cramps with menses
- 63. abdominal pain is sharp and steady
- 64. vomiting, 4-5 times this a.m.

#### ADDITIONS TO PROBLEM FORMULATION SHEETS

### Film 8

The attached sheet lists the cues from film 8 which the physicians used to generate problem formulations and diagnostic possibilities.

You are to use this sheet to do the following:

- 1. In reading the attached list, you may notice some cues which you did not record, but which you now consider to be relevant to one or more of your problem formulations (or diagnostic possibilities). IF THIS IS THE CASE, add these cues to your response sheets. Record the cue number(s) in the space under the relevant problem formulation (or in the space next to the relevant diagnostic possibility).
- 2. Having read the attached sheet, you may have thought of some additional problem formulations. IF THIS IS THE CASE, record the title of each additional formulation on a response sheet. Underneath, in the space provided, list the number(s) of all relevant cues.
- 3. Having read the attached sheet, you may have thought of some <u>additional diagnostic possibilities</u>, related to the problem formulations you recorded after viewing the film, or related to the additional problem formulations (if any) you recorded in step 2 above. IF THIS IS THE CASE, record the title of each additional diagnostic possibility on the back of the appropriate problem formulation sheet. Next to it, in the space provided, list the <u>number(s)</u> of all cues of particular relevance.

Use the pen that has been provided for you to record any of the above additions in your response booklet.

When you are finished please raise your hand.

FILM 8

List of cues which the physicians used to generate problem formulations

1.	age 19
2.	female
3.	student nurse
4.	no fever
5.	obese
6.	leaning forward, clutching abdomen
7.	anxious manner & tone of voice
8.	her condition is stable at present
9.	abdominal pain
10.	sudden onset of pain this a.m.
11.	pain is diffuse, not localized
12.	steady increase of pain
13.	pain is sharp & steady, not crampy
14. 15. 16. 17. 18.	<pre>vomiting, 4-5 times this a.m. vomits thin yellow liquid no bowel movement this a.m. no recent change in bowel movements had cheese sandwich and milk at midnight, 6 hours before onset of abdominal pain &amp; vomiting</pre>
19.	has been feeling "funny" recently
20.	has been "tired & edgy" recently
21.	admits anxiety about pregnancy
22.	last menses 8 weeks ago
23.	sexually active
24.	has broken off with boyfriend 1 month ago
25.	he was using contraception; she was not
26. 27.	<pre>stopped taking birth control pills l year ago, due   to vaginal itch vaginal itch re-occurred 2 days ago</pre>
28.	increased thirst recently
29.	increased intake of fluids recently
30.	increased appetite recently
31.	no recent weight change
32.	increased frequency of urination recently

APPENDIX E

.

•

.

### QUESTIONNAIRE

The questionnaire was the same for both treatment groups except for items 14 and 15. The copy of the questionnaire contained in this appendix includes the Treatment II version of these items. The Treatment I version of these items was as follows:

- 14. The second viewing of the film helped me to consolidate my understanding of the case.
- 15. The second viewing of the film was not worthwhile.

# QUESTIONNAIRE

-1-

## <u>Part 1</u>

Please read carefully each of the statements below. For each statement, indicate your opinion by circling <u>one</u> of the five response options:

SA = strongly agree A = agree NO = no opinion D = disagree SD = strongly disagree

Statements

Response Options

1.	The instructions were generally clear						
	and easy to follow	•	SA	Α	NO	D	SD
2.	The instructional sessions were too long .	•	SA	Α	NO	D	SD
3.	The actors who played the role of the patients in the films were very convincing	•	SA	A	NO	D	SD
4.	The films provided a realistic simulation of the early part of the clinical workup .	•	SA	A	NO	D	SD
5.	The dialogue in the films was sometimes difficult to follow	•	SA	A	NO	D	SD
6.	The physicians in the films did a good job of interviewing the patients	•	SA	A	110	D	SD
7.	I enjoyed watching the films	•	SA	Α	NO	D	SD
8.	As I watched the films, I was able to put myself into the role of the doctor	•	SA	A	NO	D	SD
9.	The films presented a good selection of medical cases	•	SA	A	NO	D	SD
10.	The feedback materials were well organized and easy to follow	•	SA	A	NO	D	SD
11.	The feedback materials were sometimes overly redundant	•	SA	A	NO	D	SD
12.	The opportunity to compare my problem formulations to those of experienced physicians helped to improve my skill in generating initial problem formulations	•	SA	A	NO	D	SD

.

# QUESTIONNAIRE (cont.)

# Statements

# Response Options

13.	I found the feedback materials interesting	SA	A	NO	D	SD
14.	The second version of the films, which portrays the physician "thinking aloud", provided me with an understanding of the <u>process</u> by which experienced physicians generate initial problem formulations	SA	A	NO	D	SD
15.	The "think aloud" segments in the second version of the films tended to disrupt my own thinking process	SA	A	NO	D	SD
16.	The self-evaluation checklists helped me to evaluate my own performance as compared to that of the experienced physicians	SA	A	NO	D	SD
17.	My ability to generate a set of initial problem formulations has improved as a result of utilizing this instructional package.	SA	A	NO	D	SD
18.	This instructional package is not appro- priate for second-year medical students	SA	A	NO	D	SD
	a. It would be more appropriate for first-year students	SA	A	NO	D	SD
	b. It would be more appropriate for third-year students	SA	A	NO	D	SD
19.	For some of the cases, I didn't have sufficient medical knowledge to be able to generate appropriate problem formulations	SA	A	NO	D	SD
<b>2</b> 0.	If a library of films like these, with accompanying feedback materials, was available to medical students, I would make use of it	SA	A	Ю	D	SD
21.	It would be more interesting to use the films and feedback materials in a group discussion setting (e.g., focal problems class) than in an individual self-					
	instructional format	SA	Α	NO	D	SD

-2-

# QUESTIONNAIRE (cont.)

### Part II

After participating in one of the previous sessions, you may have pursued your interest in one or more of the cases outside of the instructional sessions. For example, you may have discussed the cases with other students or a faculty member, or you may have looked up materials pertaining to the cases in a medical reference book. Please indicate below the ways (<u>if any</u>) in which you pursued your interest in the cases outside of the instructional sessions. Check all items that are applicable.

 Case 1 (a 21-year-old college senior who complains of fatigue, abdominal pain and diarrhea). a. I discussed the case with other student(s). b. I discussed the case with faculty member(s). c. I looked up relevant reference materials. d. other (specify)
 Case 2 (a 43-year-old landlady of a boarding house who complains of chest pain). a. I discussed the case with other student(s). b. I discussed the case with faculty member(s). c. I looked up relevant reference materials. d. other (specify)
 Case 3 (a 30-year-old taxi driver who complains of urinary distress). a. I discussed the case with other student(s). b. I discussed the case with faculty member(s). c. I looked up relevant reference materials. d. other (specify)
 Case 4 (a 40-year-old carpenter who complains of chest pain incurred while wrestling). a. I discussed the case with other student(s). b. I discussed the case with faculty member(s). c. I looked up relevant reference materials. d. other (specify)
 Case 5 (a 19-year-old college sophomore who complains of headache and sleepiness). a. I discussed the case with other student(s). b. I discussed the case with faculty member(s). c. I looked up relevant reference materials. d. other (specify)
 Case 6 (a 29-year-old lawyer who complains of low back pain) a. I discussed the case with other student(s). b. I discussed the case with faculty member(s). c. I looked up relevant reference materials. d. other (specify)

# QUESTIONNAIRE (cont.)

-4-

### Part III

Please indicate below any comments regarding this instructional package, or any suggestions for the use of these materials with medical students.

### Part IV

We are interested in knowing how much contact with actual patients you have had prior to participating in this experiment. Please list below any type of experience you have had that involved contact with patients (e.g., experience as a physician's assistant, as a nurse, as a paramedical assistant). For each type of experience, please indicate the extent of the experience (e.g., 20 hours/week for 6 weeks; 40 hours/week for 1 year).

Type of Experience

.

Extent of Experience

274

APPENDIX F

.

,

,

#### PF AND R-PF SCORES: INSTRUCTIONS

#### A. PF score

- 1. This score is based on the titles of the S's problem formulations (PF) and diagnostic possibilities (DP).
- 2. Each PF or DP title is scored as follows:
  - a. If the S's title is equivalent to one of the titles on the scoring sheet:
    - (1) under the column "Type of Response," circle the code(s) corresponding to the way(s) in which the S recorded the title.
    - (2) under the column "Pts.", circle the number of points for the title.
    - (3) if the S has recorded a title in a way for which there is no code under the column "Type of Response," write in the type of response the S gave.
    - (4) if the S fails to list any cues for a title, do not score this title.
    - (5) if the S lists more than one DP in a single response space, score only one of the DPs listed: the one with the highest number of points.
    - (6) if, on reading the S's tentative assessment, he mentions a title that was not listed on a response sheet, this title may be scored <u>providing</u> that the S mentions at least two cues that led him to consider it. Under the column "Type of Response," write "in TA," and circle the points for the title.
  - b. If the S's title is not equivalent to one of the titles in the scoring sheet:

Check the list of "Other Acceptable Responses." If the title appears in this list, write in the title, and circle one point.

3. Sum the number of points circled.

### PF and R-PF Scores: Instructions Continued

### B. <u>R-PF score</u>

- This score is based on the S's statements of functional relationships between PFs or DPs, as recorded in his tentative assessment (TA).
- 2. Each such statement is entered on the scoring sheet as follows:
  - a. If a statement is equivalent to an item on the scoring sheet, circle the number of points next to the item.
  - b. If a statement is not equivalent to an item on the scoring sheet: check the list of "Other Acceptable Responses." If the statement appears on this list, write in the statement and circle one point.
- 3. Keep in mind that statements of relationships between cues (rather than PFs or DPs) are not to be scored.
- 4. If the S lists functional relationships between PFs or DPs on his PF response sheets, but fails to mention them again in his TA, such relationships should be entered on the scoring sheet as indicated in rule B2. The scorer should indicate for such relation-ships "not in TA."
- 5. Sum the number of points circled.

#### A. CUE score

- 1. The CUE score is based on the cues the S recorded, irrespective of the PF (or DP) under which he listed them.
- 2. The entries for this score are made under the column labeled "CUE."
- 3. For each cue the S recorded, circle the number of points on the scoring sheet corresponding to the cue.
- 4. Sum the number of points circled in the CUE column.

#### B. CUE-PF score

- 1. The CUE-PF score is based on the cues the S recorded under PF (or DP) titles included in each category across the top of the scoring grid.
- 2. The entries for this score are made in the cells of the cue (rows) x category (columns) scoring grid.
- 3. The PF (or DP) titles that are included in each category are specified at the end of the scoring grid.
- 4. For each PF (or DP) title the S recorded:
  - a. determine if this title is included in one of the scoring categories;
  - b. if so, circle the number of points (in the appropriate category x cue cell) for each cue the S recorded under this title.
  - c. if there is an (\*) next to a cue, this indicates that the S must mark the cue as "negative" for PFs in that column. If the S did not do so, change the sign of the cue from + to -, and circle the points.
- 5. After completion of step 4, sum the points circled (across columns) in each row and enter this sum in the column "Tot."
- 6. Sum the points recorded in the "Tot." column.

#### CUE and CUE-PF Scores: Instructions Continued

### C. Both scores

1. In order for a cue to be scored, it is <u>not</u> necessary that the S mention all the details included in the cue description on the scoring sheet.

Example: Film 7 If the S lists "fever 101-102," or "fever 2 days," or "fever," he gets credit for cue 1.

However, it is necessary that the S list sufficient detail for the scorer to be able to determine unequivocally the cue to which the S's response refers.

#### Example: Film 7

If the S lists "cough," it is impossible to determine whether he is referring to cue 2, 3, 4 or 5. Thus, no points are scored.

2. If the S recorded a cue that is <u>clearly incorrect</u> (e.g., film 7: "no fever," "weight <u>gain</u>"), write the letters "inc" in the cell for the cue, and change the sign in front of the points from (+) to (-), or if the sign is already (-), leave it as (-).

1

I

Subject\_\_\_\_\_ PF score\_\_\_\_\_ R-PF score\_\_\_\_\_

### PF Scoring Key--Film 7

#### Titles Type of Response Pts. 1. cough and/or fever $\mathbf{PF}$ 0 2. acute respiratory PF DP under 1, 3, 6 6 infection 3. respiratory problem $\mathbf{PF}$ DP under 1 3 DP under 1, 3 4. chronic respiratory $\mathbf{PF}$ 6 problem 5. cancer (lung/ $\mathbf{PF}$ DP under 1, 3, 46 bronchogenic) 6. infection $\mathbf{PF}$ DP under 1, 3 3 7. acute bronchitis DP under 1, 2, 3, 6, 9 3 8. pneumonia DP under 1, 2, 3, 6 3 9. bronchitis $\mathbf{PF}$ DP under 1, 2, 3, 4, 6 3 10. C.O.L.D./emphysema $\mathbf{PF}$ DP under 1, 3, 4 6 11. chronic bronchitis $\mathbf{PF}$ DP under 1, 3, 4, 9 3 12. TB DP under 1, 2, 3, 4, 6 3 13. SVC syndrome $\mathbf{PF}$ DP under 5 3 14. other PF (maximum of 3) $\mathbf{PF}$ a.\_\_\_\_\_ 1 $\mathbf{PF}$ 1 b.\_\_\_\_\_ $\mathbf{PF}$ с. 1 15. other DP (maximum of 3) DP under 1 a. \_\_\_\_\_ b.\_\_\_\_\_ 1 1 с.

# R-PF Scoring Key--Film 7

	Statements of relationships	Pts.	
1.	acute respiratory infection superimposed on chronic respiratory problem or cancer	6	
2.	SVC syndrome 2 <sup>e</sup> to cancer	3	
3.	other (maximum of 3)		
	a	1	
	b	1	
	C	1	

Subject\_\_\_\_\_ CUE Score\_\_\_\_\_ CUE-PF Score

	والمراجع والمحمد والمرجع والمحمد والمرجع والمرجع والمحمد والمحمد والمحمد والمحمد والمحمد والمحمد والمحمد والمح	T	1				
		CUE		C	UE-PF		
			ARI	CRP	CA	RP	TOT
1.	fever 101-102 (2 days)	3	+3	0	0	0	
2.	chronic a.m. cough (2-3 yrs.)	3	0	+3	+3	+3	
З.	cough worse (2 days), acute cough	3	+3	0	+3	+3	
4.	<pre>woke up coughing (2 nights ago); got back to sleep propped on 2 pillows, orthopnea</pre>	2	+2	0	0	+2	
5.	wheezy cough observed	1	+1	+1	+1	+1	
6.	sputum with a.m. cough (2-3 yrs.), productive a.m. cough	3	-3	+3	+3	+3	
7.	change sputum color (yellow to green, 2 days)	3	+3	0	0	+3	
8.	blood flecks in sputum (2 days), hemoptysis	2	+2	0	+2	+2	
9.	weakness in leg muscles climbing stairs (3 wks.)	2	0	+2	+2	+2	
10.	weight loss (5-10 lbs., in 1 yr.)	3	- 3	+3	+3	0	
11.	smoker (20+ pack yrs.)	3	0	+3	+3	+3	
12.	wife noticed face is "ruddier and chubbier" recently	2	-2	+2	+2	+2	
13.	trouble buttoning collar recently, neck enlargement	3	- 3	0	+3	+3	
14.	age 57	2	-2	+2	+2	0	
15.	male	2	<b>-</b> 2	+2	+2	0	
16.	executive	1	-1	0	0	0	
17.	no chills	1	+1*	-1	-1	0	
18.	no temp. spikes	1	+1*	-1	-1	0	
19.	no s.O.B. on exerticn	1	0	+1*	-1	+1	
20.	physical and lab/X-ray (OK, 9 mo. ago)	1	-1	+1*	0	0	

CUE and CUE-PF Scoring Keys: Film 7

ARI = acute respiratory infection (pneumonia, acute bronchitis, viral/ strep/bacterial/pneumococcal/infection, pneumonitis,\_\_\_\_\_)

CRP = chronic respiratory problem (C.O.L.D./C.O.P.D., chronic bronchitis, bronchiectasis,\_\_\_\_\_)

CA = cancer (brochogenic/lung), including S.V.C. syndrome 2<sup>e</sup> to cancer, and metastases to thyroid

RF = respiratory problem (Code <u>only</u> those cues <u>not</u> coded under ARI, CEP, or CA)

Subject\_\_\_\_\_ PF score\_\_\_\_\_ R-PF score\_\_\_\_\_

# PF Scoring Key--Film 8

	Titles	Type	e of Response	Pts.
1.	abdominal pain and/or vomiting/abdominal problem	PF		3
2.	GI problem	$\mathbf{PF}$	DP under l	6
3.	appendicitis		DP under 1, 2	6
4.	gall bladder		DP under 1, 2	3
5.	gastroenteritis		DP under 1, 2	3
6.	ulcer		DP under 1, 2	3
7.	pancreatitis		DP under 1, 2	3
8.	polydypsia and/or polyuria	$\mathbf{PF}$		0
9.	diabetes mellitus	$\mathbf{PF}$	DP under 1, 8	6
10.	psychological	PF	DP under 1, 2, 8, 11, 12	6
11.	amenorrhea/dysmenorrhea	$\mathbf{PF}$		0
12.	pregnancy	$\mathbf{PF}$	DP under 1, 11	6
13.	ectopic pregnancy	-	DP under 1, 11, 12	6
14.	GU infection/problem	$\mathbf{PF}$	DP under 8, 11, 12	6
15.	vaginal itch	$\mathbf{PF}$		0
16.	vaginitis/monilial	$\mathbf{PF}$	DP under 12, 14, 15	3
17.	cystitis/urinary tract infection	PF	DP under 1, 8, 14, 15	3
18.	V.D.		DP under 14, 15, 16, 17	3
19.	P.I.D.		DP under 1, 14, 15	3
20.	obesity	$\mathbf{PF}$		3
21.	other PF (maximum of 3)			
	a	$\mathbf{PF}$		1
	b	PF		1
	c	$\mathbf{PF}$		1
		,	•	

·

PF\_Scoring Key--Film 8 Continued

Titles	Type of Response	Pts.
<pre>22. other DP (maximum of 3)     a b c</pre>	DP under	1 1 1

# R-PF Scoring Key--Film 8

	Statements of relationships	Pts.
1.	psychological problem aggravating organic problem(s)/organic problem(s) with "psychologi- cal overlay"	3
2.	diabetes predisposes to GU infection/cystitis/ vaginitis	3
3.	pregnancy: increases likelihood of GU infection/ vaginitis	3
4.	other (maximum of 3)	
	a	l
	b	1
	C	1

		CUE					CUE-PI	Du .				
			Pre	. d.							G.U.	
	CUE		Nor.	Ect.	Psych.	Gen.	App.	G.B.	G.E.	D.M.	inf.	Tot
1. aye 19		г	+1	0	0	0				0	0	
2. famale		5	+2.	0	0	0				0	+2	
3. studen	t nurse	-	0	0	+1	0				-1	0	
4. no fevu	er, T=98.8		0	+1*	-1	0	+1*			-1	+1*	
5. obese		2	-2	-2	+2 -	0		+2		+2	-2	
6. leanin abdo	g forward, clutching men	2	0	+2	0	+2				0	0	
7. anxiou of v	s manner & tone oice	2	+2	0	+2	0				-2	-2	
8. her co	ndition is stable at sent	ч	-1	+1*	-1	0	+1*			1	- 1	
9. abdomi	nal pain	m	0	+3	0	+3				+3	+3	
0. sudden a.m.	onset of pain this	5	-2	+2	0	+2				-2	0	
l. pain d all	iffuse, not localized, over"	m	۳ ۱	+3*	0	+3	+3*			0	0	
2. steady	increase of pain	н	-1	+1	-1	+1				-	0	
3. pain s not	evere, sharp & steady, crampy	2	-2	+2	-2	+2			+2*	0	°.	
4. vomiti a.m.,	ng, 4-5 times this "morning sickness"	e	£+	0	0	+3		+3	+3	- m	е-	
5. vomits	thin yellow liquid	-	0	7	-1-	1+		+1	1+	1-	-1	
.6. no bow	el movement this a.m.	-	-1-	7	-1-	1+			+1*	-1	-1	
7. no rec move	ent change in bowel	2	-2	-2	-2	+2			+2*	-2	-2	
.8. cheese 12 a of p	sandwich and milk at	г	- 1	7	-1	I+.				7	-1	

		ot.														
	. П.	lf. T		0		0	5	, N	0	5	ņ	, m	5		ņ	
	ບ	<u>ب</u> نا ۲					+			+	+			-	+	
		D.M	+1	+2	n I	0	-2	-2	- 2	0	÷	+ +	+2	Ŧ	÷	0
	G.I.	G.E.														
		G.B.														
		.dqA														
		Gen.	- 1	0	۲ ۲	- 3	- 2	- 2	2 	- 2	۳ ا	0	0	0	۳ ۱	ŗ
		Psych.	+1	+2	+3	+3	0	+2	0	0	o	+3	+2	0	0	r,
	:rej.	Ect	J	0	0	0	0	1	C	- 2	0	m 1	- 2	7	0	- -
		Nor.	+1	2+	1) 1)	(1) +	<7 +	C	+ 2	c	"+	0	O	7	+3	7
а Б			r-i	~	m	n	C4	(1	(1	5	m	m	2		ion 3	
		GCB.	<pre>0. feeling " int" result; "not not sail"</pre>	0. "Litea Videy" micently	l. adasta anxisto, a sat perspeciej	<ol> <li>Last menses 3 wks. ago, pariod 5 meeks late</li> </ol>	<ol> <li>sexually active</li> </ol>	<ol> <li>broke off with boyfriend ene wouth ago</li> </ol>	<ol> <li>he word contraception; she did not; ineflective contraception</li> </ol>	<ul> <li>stopped birth control pills</li> <li>(i gr. a.s) due to Vignal (tch</li> </ul>	7. vayinal itch re-occurred 2 days ayo	<ol> <li>increased thirst recently, polydipsia</li> </ol>	9. increased appetite recently, polyphagia	0. no recent weight change	<pre></pre>	<pre>2. physical/lab/Xray OK     (1 yr. ago)</pre>

CUE and CUE-PF Scoring Sources Film 8 Continued

CUE and CUE-PF Scoring Keys: Film 8 Continued

1

Ect. = ectopic pregnancy Nor. = normal unspecified Preg. = pregnancy

Psycho . = psychological/anxiety/depression

GI = gastrointestinal

Gen.: score cues listed under GI problem (unspecified) and/or any of the following GI problems:

appendicitis pancreatitis gall bladder food poisoning gastroenteritis under the columns labeled "app.," "G.B.," "G.E.": If the S lists any of the cues in these columns, score such cues in the column GI "Gen." and in these columns.

D.M. = diabetes mellitus

G.U. inf. = genito-urinary infection
 score cues listed under any of the following:

G.U. infection V.D. (any type) pelvic inflammatory disease cystitis

vaginitis

ļ
# APPENDIX G

•

.

# ANALYSIS OF THE STRUCTURE OF A SET OF PROBLEM FORMULATIONS

Example of two physicians' responses for Film 1

# Subject: A

	Problem Formulations	No. PF	No. S	Structural Features
1.	GI disorder la. ulcerative colitis/ileitis	1 1	1	H: l(a,b,c) C: la-lb-lc
2. 3. 4.	lb. GI malignancy lc. psychogenic diabetes anemia (2 <sup>e</sup> to la or lb) renal problem	1 1 1 1	1 1 1	S: 1,2,3,4 R: 3/la or lb
		7	4	

#### Subject: B

	Problem Formulations	No. PF	No. S	Structural Features
1. 2. 3.	intestinal inflammation la. regional enteritis lb. ulcerative colitis anemia (2 <sup>e</sup> to 1) cardiovascular (along with 1)	1 1 1 1	1 1 1	H: l(a,b) C: la-lb S: l,2,3 R: 2/l
	•	5	3	

Note:	No. PF = number of problem formulations No. S = number of subspaces
	Features: H = hierarchical organization C = competing formulations S = subspaces P = functional relationships
	R = functional felationships

A response was not counted as a problem formulation unless there were at least two cues associated with it.

APPENDIX H

TABLE	28	Compari	ison	of	the	e Par	ticipants	with	the	Refusal/
	No	Contact	Grou	ip d	on E	ocal	Problems	Exam	Scoi	ces.

Group	n	Mean	St. Dev.
Participant	47	72.06	7.115
Refusal/No Contact	15	72.67	7.556

90% Confidence Interval:

 $-4.187 \leq (\mu_1 - \mu_2) \leq 2.967$ 

Note: The comparison concerns the participants in the experiment versus the students who were sampled but refused to participate (or could not be contacted) except for the four students with exam scores considerably below all other students (i.e., 20% of refusal/no contact group, or 6% of target population). Scores were not available for one participant and for one refusal.

Variable	Treatment I	Treatment II	Control
No. Subspaces	8.3	7.9	6.5
No. Problem Formulations	15.0	12.3	9.5

TABLE 29.--Adjusted Means on Number of Subspaces, and Number of Problem Formulations, by Experimental Condition.

TABLE 30.--Multivariate Analysis of Covariance on Number of Subspaces and Number of Problem Formulations.

F Tests	df	F	p		
Multivariate F test	4,86	5.0314	.0011		
Stepdown F tests					
no number of problem formulations	2,44	6.2399	.0042		
no number of subspaces	2,44	4.0060	.0254		

Variable	Comparison	Confidence Interval	(l-a)%
No. Subspaces	T1-T2 0.4	± 1.39	95%
	T1-C 1.8*	± 1.41	95%
	T2-C 1.4	± 1.39	95%
No. Problem			
Formulations	T1-T2 2.8	± 3.12	95%
	T1-C 5.5*	± 5.03	99.9%
	T2-C 2.8	± 3.11	95%

TABLE 31.--Scheffé Post Hoc Comparisons on Number of Subspaces and Number of Problem Formulations.

Note: T1 = Treatment I; T2 = Treatment II; C = Control.

\*Significant group differences at the .05 level of better.

APPENDIX I

.

### HOMOGENEITY OF REGRESSION

The sample estimates of the population regression slopes, reported in Table 32, were sufficiently dissimilar to indicate that a test of the ANCOVA model assumption of homogeneity of regression should be conducted. For each dependent variable, a procedure described in Winer (1962) was used to test the null hypothesis that the population regression coefficients for the three experimental conditions are equal (H<sub>0</sub>:  $\beta_1 = \beta_2 = \beta_3$ ). The test consists of: (1) partitioning the within-group variance into two components (S1 and S2), and (2) calculation of an F ratio (F = S2/S1) which, if significant, leads to rejection of the null hypothesis. Since the desired outcome in conducting this test is to fail to reject the null hypothesis, the probability of a Type I error ( $\alpha$ ) for the set of four tests was set at .20 (i.e.,  $\alpha$ =.05 for each test) in order to reduce the possibility of making a Type II error. The results of the tests of homogeneity of regression of each dependent variable (CUE, PF, CUE-PF, R-PF) on the covariate (Focal Problems Exam) are reported in Table 33. As indicated in this table, the F tests were nonsignificant for PF, CUE-PF and R-PF indicating that for these three variables there was no evidence of departure from homogeneity of regression. For the variable CUE, on the other hand, the F test was significant. However, upon closer examination of the data,

it was found that there was some evidence to suggest that this outcome constituted a Type I error (i.e., rejection of the null hypotheses when it is in fact true).

Examination of the regression and correlation coefficients for the covariate and CUE (Table 32) indicated that the significant F test had occurred due to the disparity between the Treatment II and control coefficients (correlations of -.32 and +.53, respectively). Since a negative correlation was guite unexpected on a priori grounds (i.e., the covariate might fail to correlate at all with the dependent variables, but should not correlate negatively), it was decided to plot the bivariate distribution for the Treatment II group (see Figure 4). The plot of this distribution revealed that 15 of the 16 data points were scattered in a random appearing cluster about the bivariate mean, but that one data point fell, at a distance from this cluster, in the corner of lower-right quadrant (i.e., the subject with the highest score on the covariate had the lowest score on CUE). The plot of the data suggested that this one extremely deviant observation was respondible for the negative coefficient for Treatment II which, although not significantly different from zero in itself, was sufficiently disparate from the control group coefficient to produce the significant F ratio in the test of homogeneity of regression. When the regression coefficients for Treatment II were recomputed with this one subject dropped from the sample,

a coefficient of zero was found for the regression of CUE on the covariate (see Table 34). With a zero coefficient for Treatment II, the disparity between the coefficients for the three groups is considerably reduced (i.e., .08, -.01, .53) and, it is reasonable to assume, (given the outcome of the test for PF based on a highly similar set of coefficients) would have led to a nonsignificant F test. It may also be noted that with this one subject eliminated, the disparity between the regression coefficients for the other three variables either remained the same (in the case of R-PF) or was reduced (in the cases of PF and CUE-PF). In sum, a careful examination of the data suggested that a highly deviant performance on the part of a single subject led to a spurious negative correlation between the covariate and CUE for Treatment II, and that this in turn was responsible for a Type I error on the test of homogeneity of regression of CUE. Thus, in the author's opinion, there was not sufficient evidence of departure from homogeneity of regression to invalidate the use of the ANCOVA model.

It should be noted that although the analyses described in this Appendix lead to the conclusion that the population regression slopes do not differ, the possibility must be borne in mind that a Type II error (i.e., failure to reject a false null hypothesis) could have occurred. If this were the case, the major conclusion of Chapter V-namely, that there was a significant treatment main effect

on the variable PF--would have to be qualified. A zero (or nearly zero) regression slope for the two treatment conditions and a moderate positive slope for the control condition (see Table 32) would suggest an ability by treatment interaction, which, given the size of the differences between the group means (reported in Table 13), would probably be ordinal in nature. If this type of interaction did exist, it would imply that although all second-year students may improve their skills in generating initial problem formulations as a result of the training, lower ability students would tend to improve to a greater degree than higher ability students.

TABLE 32.--Regression and Correlation Coefficients for the Covariate with each Dependent Variable, by Experimental Condition.

Dep. Variable	Treatment I	Treatment II	Control
CUE	b = .090	b =485	b = .834*
	r = .089	r =317	r = .533*
PF	b = .135	b =014	b = 1.073*
	r = .084	r =011	r = .570*
CUE-PF	b = .296	b =455	b = 1.673*
	r = .168	r =154	r = .505*
R-PF	b = .244*	b =041	b = .320*
	r = .525*	r =075	r = .504*

\*Coefficient is significantly different from zero at p < .05.

Dependent Variable	Sources Variatio	of on	đf	SS	F
CUE	Subjects:	Group S1 S2	44 42 2	4123.24 3506.45 616.79	3.6939*
PF	Subjects:	Group S1 S2	44 42 2	5779.84 5247.40 532.44	2.1308
CUE-PF	Subjects:	Group S1 S2	44 42 2	15796.22 14157.08 1639.14	2.4314
R-PF	Subjects:	Group S1 S2	44 42 2	601.92 555.80 46.12	1.7426

TABLE 33.--Tests of Homogeneity of Regression for CUE, PF, CUE-PF, R-PF.

\*Significant at p < .05.



Figure 4. Bivariate Distribution of Treatment II scores on the Covariate and the Dependent Variable CUE.

TABLE 34.--Regression and Correlation Coefficients for Treatment II with one Deviant Subject Eliminated.

Dependent Variable	b	r	
CUE	018	013	
PF	.216	.150	
CUE-PF	.371	.129	
R-PF	045	072	

APPENDIX J

### MODIFICATIONS OF THE OUTCOME FEEDBACK VERSION OF THE TRAINING MATERIALS

## Recommendations for Instructional Application with Second-Year Medical Students

It is believed that the following modifications of the outcome feedback version of the training materials would result in an improved instructional package for training second-year medical students in the generation of initial problem formulations.

1. Modification of the instructional sequence with respect to the second viewing of the film.

a. In the present version of the materials, the film of the doctor-patient interview is presented a second time after the student receives the first feedback sheet and before he receives the second feedback sheet. This interpolation of the film between the two feedback sheets was intended to reduce the possibility of information overload and/or attentional decrement which, it was believed, might occur if all of the feedback material was presented at the same time. However, a number of students commented (in section 3 of the questionnaire) that it would be preferable to view the film after having received all the feedback material, particularly since it was the problem formulations on the second sheet which they most often failed to generate. In retrospect, the experimenter believes that precautions against information overload and/or attentional decrement

probably were not needed, and that presentation of the film after all of the feedback information would probably enhance the student's ability to assimilate this information.

b. Treatment I group responses to items 14 and 15, section 1 of the questionnaire, indicated that many students did not find the second viewing of the film to be worthwhile. Thus, a second recommendation is that this viewing be made <u>optional</u>. In a self-instructional setting this would mean that the student could elect not to view the film again at all, or, for an exceptionally difficult case, could chose to view it again several times.

2. Modification of the response format.

In the present version of the materials, a sizable amount of time was required for the student to record his responses (i.e., problem formulations, plus a cue list for each formulation). It also required a good deal of time for the experimenter (and would require for the instructor) to read and score the students' responses. The following modifications of the response format would considerably reduce both of these time requirements:

a. The student lists all cues he obtained while viewing the film (i.e., data he used to generate one or more problem formulations).

b. The student receives a "Cue Feedback Sheet." This sheet presents anumbered list of all cues the physicians used to generate problem formulations.

c. The student records each problem formulation he generated, but rather than <u>writing out</u> a cue list for each formulation, he records the <u>numbers</u> of the cues on the Cue Feedback Sheet that are relevant to each of his formulations.

This response format was originally considered in designing the training experiment but was rejected because it would permit the student to generate problem formulations (at least in part) on the basis of the Cue Feedback Sheet, rather than solely on the basis of his own naturalistic observation while viewing the film. However, since the results of the experiment indicated that, at least for second-year medical students, skill in cue detection on the basis of naturalistic observation was already well established, it is believed that the proposed modification of the response format would not diminish the instructional value of the exercises.

To summarize: a <u>revised</u> <u>instructional</u> <u>sequence</u>, incorporating the above modifications, would include the following steps:

- 1. The student reads the "nurse's sheet."
- The film of the doctor-patient interview is presented.
- 3. The student records his cue list.
- The student receives a "Cue Feedback Sheet" (described above).

- 5. The student records his problem formulations, and, for each formulation, lists the numbers of the cues (on the Cue Feedback Sheet) of relevance it. He then writes his tentative assessment.
- 6. The student receives a "Problem Formulations Feedback Sheet" (i.e., a composite of the two feedback sheets provided in the present study).
- 7. The film is presented a second time (optional).
- The student completes the Self-evaluation Checklist.

